现代数学丛书

陈翰馥　朱允民　著

随机逼近

STOCHASTIC APPROXIMATION

CHEN HANFU
ZHU YUNMIN

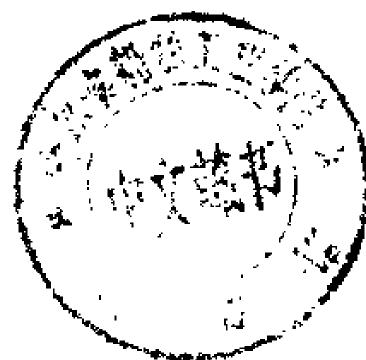
上海科学技术出版社

391998

·现代数学丛书·

随 机 逼 近

陈翰馥 朱允民 著



上海科学技术出版社

·现代数学丛书·
随 机 逼 近
陈翰馥 朱允民 著
上海科学技术出版社出版、发行
(上海瑞金二路 450 号)
新华书店上海发行所经销 商务印书馆上海印刷厂印刷
开本 787×1092 小 1/16 印张 18.5 插页 4 字数 230,000
1996 年 10 月第 1 版 1996 年 10 月第 1 次印刷
印数 1—1,200
ISBN 7-5323-3906-8/O·191
定价：26.00 元

内 容 提 要

寻找带误差地量测到的未知回归函数的零点或极值，是系统辨识、适应控制、模式识别、适应滤波和神经元网络等领域中都要遇到的问题．随机逼近提供了解决这一问题的递推方法．本书在一些准备材料之后，在第2章中叙述了分析随机逼近算法常用的三种基本方法，侧重解释这些方法本身，而不在于为收敛性所加条件的强弱．第3章介绍随机变界截尾的随机逼近方法，这是一种修改了的 Robbins-Monro 算法，后者只适用较窄的一类回归函数和量测噪声．这一章还分析了算法的收敛性、收敛速度及稳健性．第4章叙述了算法的渐近正态性，算法中用慢衰减增益时平均估计的渐近有效性．第5章讨论连续时间算法及相应的性质．最后三章涉及随机逼近的应用．第6章应用随机逼近收敛性定理解决动态系统中各种参数估计问题．第7章讨论信号处理中的适应滤波问题，表明这类问题可以用随机逼近来解决．最后一章，用随机逼近的结果和处理方法来研究随机并行处理及多传感器数据融合问题．

本书适用于从事随机优化、系统与控制、数理统计及信号处理等领域的学生及科研人员，对从事实际优化问题的人员，本书也是一本有价值的参考书．

D369/02

Modern Mathematics Series

STOCHASTIC APPROXIMATION

Chen Hanfu   Zhu Yunmin

Shanghai Scientific & Technical Publishers

Stochastic Approximation

Abstract

The problem of searching for zeros or extrema of a unknown regression function observed with noise is often met in various fields including system parameter identification, adaptive control, pattern recognition, adaptive filtering and neural networks. Stochastic approximation provides a recursive way to solve this problem. Following some preparatory material the book describes three basic methods in Chapter 2 which are commonly used for analyzing approximation algorithms with emphasis on explanation of the methods themselves rather than on conditions required for convergence. Chapter 3 presents the stochastic approximation algorithm truncated at randomly varying bounds, a modification to the classical Robbins-Monro algorithm, which is applicable for a rather restrictive class of regression functions and observation noises. Convergence, convergence rates and robustness of the algorithm proposed are also analyzed in this chapter. Chapter 4 demonstrates the asymptotic normality of the algorithm and asymptotic efficiency of the averaged estimate when a slowly decreasing gain is applied. Continuous-time algorithms and their corresponding properties are presented in Chapter 5. The last three chapters are of application oriented. Convergence theorems of stochastic approximation are applied to various parameter estimation problems in dynamic systems. This is done in Chapter 6. Adaptive filtering in signal processing is dealt with in Chapter 7. It is shown that the problem in question can be solved by stochastic approxi-

mation. In the last chapter the parallel processing and multi-sensor data fusion are considered with the help of results and techniques developed for stochastic approximation.

The book is designed for students and researchers in stochastic optimization, system and control, mathematical statistics and signal processing. Those who work in practical optimization problems will also find the book a valuable resource.

《现代数学丛书》编辑委员会

名誉主编　苏步青

主　　编　谷超豪

委　　员　（以姓氏笔划为序）

丁夏畦　王梓坤　叶彦谦

石钟慈　冯克勤　刘应明

严志达　杨　乐　吴　方

李大潜　陈希孺　陈翰馥

张恭庆　胡和生　姜伯驹

梁友栋　曹锡华　程民德

Modern Mathematics Series
Editorial Committee

Honorary Editor-in-Chief Su Buchin

Editor-in-Chief Gu Chaohao

Members

Cao Xihua	Chen Hanfu
Chen Xiru	Cheng Minde
Ding Xiaqi	Feng Keqin
Hu Hesheng	Jiang Boju
Li Tatsien	Liang Youdong
Liu Yingming	Shi Zhongci
Wang Zikun	Wu Fang
Yan Zhida	Yang Le
Ye Yanqian	Zhang Gongqing

出版说明

从60年代起，由华罗庚教授任主编的《现代数学丛书》编辑委员会曾组织编著，并由我社出版了多部具有很高水平的数学学术专著，有几部专著已在国外出了外文版，受到国内外数学界和广大读者的高度重视，获得了很高的评价．原编委会中华罗庚、关肇直、吴新谋3位教授虽已先后逝世，但他们为本《丛书》所作出的贡献迄今仍为人们所敬仰、怀念．由于某些客观原因，《现代数学丛书》的出版工作曾一度停顿．

为了适应现代数学的迅速发展，更好地反映我国数学家近几年的优秀研究成果，必须大力加强《现代数学丛书》的规划、编辑、出版工作，充实编委会的力量．考虑到不少编委年事已高，经向原编委会中大部分同志及数学界有关专家广泛征求意见后，于1990年对编委会作了调整，补充了一些著名的中年数学家和学科带头人，建立了新的编委会，并进一步明确了本丛书的宗旨．

《现代数学丛书》新的编辑委员会由苏步青教授任名誉主编、谷超豪教授任主编，18位著名数学家任委员．编委会负责推荐（或审定）选题和作者，主持书稿的审核等工作．

《现代数学丛书》的宗旨是：向国内外介绍我国比较成熟的、对学科发展方向有引导作用的、国内第一流水平的数学研究成果，反映我国数学研究的特色和优势，扩大我国数学研究成果的影响，促进学科的发展和国内外的学术交流．

为了实现上述宗旨，本丛书将陆续组织出版在基础数学、应用数学和计算数学方面处于学科发展前沿、有创见且具有系统完整

研究成果的现代数学学术专著.

为出版好《现代数学丛书》，我们热切地期望着数学界各位专家的大力支持和悉心指导，并欢迎广大读者提出宝贵的建议和意见.

上海科学技术出版社

序　言

在许多理论课题或实际应用中，经常要求一个函数的零点或极值，如果函数有已知的解析表达式，那么在理论上解决这个问题并不困难；如果虽不知函数的表达式，但它在任一点的取值可以无误差地量测到，那么有不少行之有效的数值方法可供选用；当既不知函数的表达式，又不能无误差地量测到函数值时，如何求函数的零点或极值，正是随机逼近要解决的问题.

不同领域的许多实际问题：例如信号处理领域中的适应波束形成，适应控制中极小化性能指标问题，模式识别统计方法中建立分割超平面问题，参数估计中的递推方法等都可归结为随机逼近问题.它在适应控制、参数辨识、随机优化等领域中起到重要作用，并在智能控制、学习系统、模式识别、神经元网络等领域中也正在显示出它的重要意义.

随机逼近创始于50年代初，Robbins-Monro[78]首先提出了求未知函数$f(\cdot)$零点x^0的一个递推算法.若在n时刻对x^0的估计为x_n，在下一时刻$n+1$，在x_n处对$f(x_n)$进行量测，但量测量$y_{n+1}=f(x_n)+\varepsilon_{n+1}$带有误差$\varepsilon_{n+1}$.在文献[78]中建议用$a_ny_{n+1}$来修正$x_n$，得到下一步时刻对$x^0$的估计值$x_{n+1}$. a_n叫增益系数.当$n\to\infty$时，它一方面要趋于0，以逐步排除量测误差对估计的影响；另一方面，a_n趋于0的速度也不能太快，否则可能会使估计值凝聚在某一点，而达不到x^0.

在Robbins-Monro的奠基性工作后，随机逼近取得蓬勃发展.主要的研究目标是考察各种相关的量测噪声[32]及拓广可适用的回归函数范围[30]，而收敛类型，除了最初的均方收敛，更多地研究概率1收敛以及弱收敛.从研究方法讲，文献[74]是用鞅收

敛方法研究随机逼近算法概率 1 收敛的有代表性的专著，基本特点是讨论独立噪声，它对 70 年代以前工作做了一个很好的总结. 之后，文献[65]把递推算法的收敛性和微分方程的稳定性建立了联系，从而使所考察的量测误差的范围有实质性扩大. 这种思想在专著[49]中得到清晰的阐明. 但这种方法要事先假定算法给出的估计值有界.

人们自然希望，既把所讨论的噪声和回归函数的范围扩大，但又不用事先假定算法有界. 另外，微分方程方法虽然富有启发性，但毕竟要从离散算法内插成连续时间的方程，然后再回到离散时间，总有绕圈子之嫌，这就是作者撰写本书的出发点之一. 我们将着重研究稍作修改的 Robbins-Monro 算法，实质性地放弃对回归函数及量测噪声的要求，同时也不必事先假定算法有界. 此外，我们将利用 Lyapunov 函数的性质，直接在离散时间上解决收敛性问题，而不再到连续时间里去“转”一圈.

本书第 1 章介绍了要用到的概率及一些准备知识，有一定专业基础的读者，可以跳过这一章. 第 2 章介绍研究随机逼近的各种方法，重在方法本身，而不拘泥于具体结果. 第 3、4、5 章研究随机变界截尾的 Robbins-Monro 算法，讨论它的收敛性、稳健性、收敛速度、渐近正态性及渐近有效性，不仅讨论离散时间，也研究连续时间算法. 求大范围极值的随机逼近，虽然很吸引人，但目前已有结果[49]，是把问题转化为连续时间的模拟退火来解决，超出了本书范围. 第 6 章讨论用随机逼近的办法来估计动态系统所含的未知参数. 第 7 章转向讨论在信号处理和通讯中有广泛应用背景的适应性滤波算法，我们将给出它的收敛性、收敛速度、稳健性、渐近正态性和渐近有效性结果. 随机并行处理和多传感器数据融合是 80 年代后期才逐渐引起国际上广泛注意的新研究方向，在第 8 章中我们将针对 Robbins-Monro 算法和适应性滤波算法给出一些有关上述研究方向的新结果.

目　　录

CONTENTS

第1章

准备知识

本章的内容包括概率论、随机微分方程及其他领域的和本书有关的一些事实及概念，主要是为了以后章节引用．叙述力求简明，基本上不加证明．不熟悉这些内容的读者，承认这些事实后，阅读本书其余章节就不会有实质困难，有兴趣的读者，可参阅文献[28, 72]．

§1.1 概率论的一些概念

设 Ω 为基本概率空间，Ω 中的点叫基本事件或样本．$\mathscr{F}$ 为 Ω 中的 σ-代数，换句话说，$\mathscr{F}$ 是 Ω 中的集合类，并满足以下条件：$\Omega\in\mathscr{F}$；若 $A\in\mathscr{F}$，则 A 的余集 $A^c\in\mathscr{F}$；$\mathscr{F}$ 对可列并封闭，即当 $A_i\in\mathscr{F}$，$i=1, 2, \cdots$时，则 $\bigcup_{i=1}^{\infty} A_i\in\mathscr{F}$．

$\mathscr{F}$ 中的元(集合)叫随机事件．定义在 $\mathscr{F}$ 上的函数 $P(\cdot)$，如果满足以下条件，则叫概率，即

1．$P(A)\geqslant 0$, $\forall A\in\mathscr{F}$；

2．$P(\Omega)=1$；

3．若 A_i 两两不相交，即 $A_i\cap A_j=\phi$, $\forall i\neq j$，那么

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i).$$

$(\Omega, \mathscr{F}, P)$叫概率空间．

定义在概率空间上的可测函数 $\xi(\omega)$，如果 $P(\omega; |\xi(\omega)|<\infty)=1$，则 $\xi(\cdot)$(或简写为 ξ)叫随机变量．

对随机变量 ξ，定义函数

$$F_\xi(x)=P(\omega: \xi(\omega)<x),\ \forall x\in R,$$

$F_\xi(x)$叫ξ的分布函数.

设ξ^i, $i=1, \cdots, l$是随机变量, 则$\xi=[\xi^1\cdots\xi^l]^\tau$叫随机矢量, ξ的分布函数是多元函数

$$F_\xi(x^1\cdots x^l)=P(\omega: \xi^1<x^1, \cdots, \xi^l<x^l).$$

如果$F_\xi(x^1\cdots x^l)$对每个自变量可微,

$$F_\xi(x^1\cdots x^l)=\int_{-\infty}^{x^1}\cdots\int_{-\infty}^{x^l} f_\xi(\lambda^1\cdots\lambda^l)d\lambda^1\cdots d\lambda^l,$$

那么$f_\xi(x^1\cdots x^l)$叫ξ的密度, 或密度函数.

如果ξ的密度为

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ \forall x\in R, \sigma>0,$$

那么称ξ的分布为正态分布, 或高斯(Gauss)分布.

n维的正态分布的密度函数为

$$\frac{1}{(2\pi)^{\frac{n}{2}}(\det R)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-\mu)^\tau R^{-1}(x-\mu)\right\},$$

$$R>0,\ x,\ \mu\in R^n.$$

现在我们来叙述数学期望的概念.

设ξ为随机变量, $P(\xi\geqslant 0)=1$, 或写作$\xi\geqslant 0$ a. s.记

$$A_{ni}=\{\omega: i2^{-n}<\xi<(i+1)2^{-n}\}.$$

ξ的数学期望$\mathrm{E}\xi$定义为

$$\mathrm{E}\xi\triangleq\int_\Omega \xi \mathrm{d}P\triangleq\lim_{n\to\infty}\left[\sum_{i=0}^{n2^n-1} i2^{-n}P(A_{ni})+nP(\xi>n)\right].$$

当$\mathrm{E}\xi<\infty$时, 称ξ可积, 对不一定非负的随机变量ξ, 可把它表为正部$\xi^+\triangleq\max(\xi, 0)$和负部$\xi^-\triangleq\max(-\xi, 0)$的差:

$$\xi=\xi^+-\xi^-.$$

如果$\mathrm{E}\xi^+<\infty$或$\mathrm{E}\xi^-<\infty$, 则定义ξ的数学期望$\mathrm{E}\xi$为

$$\mathrm{E}\xi=\mathrm{E}\xi^+-\mathrm{E}\xi^-.$$

当$\mathrm{E}\xi$有定义时, 它还可表达为 Lebesgue-Stieltjes 积分

$$\mathrm{E}\xi=\int_\Omega \xi \mathrm{d}P=\int_{-\infty}^{\infty} x\mathrm{d}F_\xi(x).$$

对服从正态分布的随机变量，$E\xi=\mu$，$E(\xi-\mu)^2=\sigma^2$. 对多维情形，$E\xi=\mu$，$E(\xi-\mu)(\xi-\mu)^\tau=R$.

设$\{\xi_n\}$为随机变量序列，ξ为随机变量.

1）如果$P(\xi_n\to\xi)=1$，则称ξ_n以概率1收敛到ξ，或ξ_n a.s.（几乎必定）收敛到ξ，记作

$$\xi_n \xrightarrow[n\to\infty]{} \xi \quad \text{a.s.}$$

2）如果对任意$\varepsilon>0$，有

$$P(|\xi_n-\xi|>\varepsilon)\xrightarrow[n\to\infty]{} 0,$$

则称ξ_n依概率收敛到ξ，记作

$$\xi_n \xrightarrow[n\to\infty]{P} \xi.$$

3）如果在$F_\xi(x)$的任意连续点上，ξ_n的分布函数$F_{\xi_n}(x)$收敛到$F_\xi(x)$，则称ξ_n依分布收敛到ξ，记作$\xi_n \xrightarrow[n\to\infty]{\omega} \xi$，或$F_{\xi_n}(x)\xrightarrow[n\to\infty]{} F_\xi(x)$.

4）如果$E|\xi_n-\xi|^2\xrightarrow[n\to\infty]{} 0$，则称$\xi_n$向$\xi$均方收敛.

$\xi_n \xrightarrow[n\to\infty]{} \xi$ a.s. 包含$\xi_n \xrightarrow[n\to\infty]{P} \xi$，而$\xi_n \xrightarrow[n\to\infty]{P} \xi$又包含$\xi_n \xrightarrow[n\to\infty]{\omega} \xi$. 均方收敛包含依概率收敛.

定理1.1.1（单调收敛定理） 如果ξ_n单调非降地收敛到ξ，并且$\xi_n\geqslant\eta$ a.s.，$E\eta^-<\infty$，那么$E\xi_n$单调非降地收敛到$E\xi$.

定理1.1.2（Fatou 引理） 设$\xi_n\geqslant\eta$（或$\xi_n\leqslant\eta$），$n=1,2,\cdots$，$E\eta^-<\infty$（$E\eta^+<\infty$），那么

$$E\liminf_{n\to\infty}\xi_n\leqslant\liminf_{n\to\infty}E\xi_n,$$

$$(\limsup_{n\to\infty}E\xi_n\leqslant E\limsup_{n\to\infty}\xi_n).$$

定理1.1.3（控制收敛定理） 设$\xi_n \xrightarrow[n\to\infty]{P} \xi$，并且存在可积控机变量$\eta$，使$|\xi_n|\leqslant\eta$ a.s. 那么$E|\xi|<\infty$，并且

$$E\xi_n\xrightarrow[n\to\infty]{} E\xi,\quad E|\xi_n-\xi|\xrightarrow[n\to\infty]{} 0.$$

在概率论中有几个常用的不等式.

Чебышев 不等式:

$$P(|\xi|\geqslant \varepsilon)\leqslant\frac{\mathrm{E}|\xi|}{\varepsilon}.$$

Jensen 不等式:

设 $\mathrm{E}|\xi|<\infty$, $g(\cdot)$为凸 Borel 可测函数, 则

$$g(\mathrm{E}\xi)\leqslant\mathrm{E}g(\xi).$$

Ляпунов 不等式:

设 $0<s<t$, 则

$$(\mathrm{E}|\xi|^s)^{\frac{1}{s}}\leqslant(\mathrm{E}|\xi|^t)^{\frac{1}{t}}.$$

Hölder 不等式:

设 $1<p<\infty$, $1<q<\infty$, $\dfrac{1}{p}+\dfrac{1}{q}=1$, 那么

$$\mathrm{E}|\xi\eta|\leqslant(\mathrm{E}|\xi|^p)^{\frac{1}{p}}(\mathrm{E}|\eta|^q)^{\frac{1}{q}}.$$

当$p=q=2$, Hölder 不等式称为 Cauchy 不等式或 Schwarz 不等式.

Minkowski 不等式:

设 $p\geqslant1$, 则

$$(\mathrm{E}|\xi+\eta|^p)^{\frac{1}{p}}\leqslant(\mathrm{E}|\xi|^p)^{\frac{1}{p}}+(\mathrm{E}|\eta|^p)^{\frac{1}{p}}.$$

C_r—不等式:

记
$$C_r=\begin{cases}1, & 若\ r\leqslant1,\\ n^{r-1}, & 若\ r>1,\end{cases}$$

则
$$\left(\sum_{i=1}^{n}|\xi_i|\right)^r\leqslant C_r\sum_{i=1}^{n}|\xi_i|^r.$$

在概率论中条件期望是一个重要概念. 设 $\mathscr{F}_1$ 是 $\mathscr{F}$ 的子 σ-代数, 即 $\mathscr{F}_1$ 本身是一个 Ω 中集合的 σ-代数, 并且任一 $A\in\mathscr{F}_1$, 必有 $A\in\mathscr{F}$.

根据 Radon-Nikodym 定理, 对任一 $\mathrm{E}\xi$ 有定义的随机变量 ξ, 必存在对 $\mathscr{F}_1$ 可测的随机变量 ζ(即对任一 $x\in R$, $(\omega:\zeta<x)\in\mathscr{F}_1$), 使对任一 $A\in\mathscr{F}_1$ 成立

$$\int_A \xi \mathrm{d}P = \int_A \zeta \mathrm{d}P \quad \text{a. s.}$$

ζ 记作 $\mathrm{E}(\xi|\mathscr{F}_1)$，称为 ξ 在 $\mathscr{F}_1$ 下的条件期望，或给定 $\mathscr{F}_1$，ξ 的条件期望.

当 $\mathscr{F}_1$ 为随机变量 η 生成的 σ-代数 $\mathscr{F}^\eta$，即包含形如 $\{\omega: \eta(\omega)<x, \forall x\in R\}$ 一切集合的最小 σ-代数，则记 $\mathrm{E}(\xi|\mathscr{F}^\eta)$ 为 $\mathrm{E}(\xi|\eta)$，称 η 条件下 ξ 的条件期望.

条件期望具有线性性质，并且 $\mathrm{E}(\mathrm{E}(\xi|\mathscr{F}_1))=\mathrm{E}\xi$，当 η 为 $\mathscr{F}_1$ 可测时，则 $\mathrm{E}(\eta\xi|\mathscr{F}_1)=\eta\mathrm{E}(\xi|\mathscr{F}_1)$，当 $\mathscr{F}_1\subset\mathscr{F}_2\subset\mathscr{F}$ 时，即 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 都为 $\mathscr{F}$ 的 σ-Z 代数，并且 $\mathscr{F}_1\subset\mathscr{F}_2$，那么

$$\mathrm{E}(\mathrm{E}(\xi|\mathscr{F}_2)|\mathscr{F}_1)=\mathrm{E}(\xi|\mathscr{F}_1).$$

当 $\xi=I_A$，即 $\xi=\begin{cases}1, & \omega\in A\\ 0, & \omega\overline{\in} A\end{cases}$ 时，则 $\mathrm{E}(I_A|\mathscr{F}_1)$ 称给定 $\mathscr{F}_1$，A 的条件概率，记作 $P(A|\mathscr{F}_1)$.

对条件期望也成立相应的单调收敛定理，Fafon 引理，控制收敛定理，以及 Чебышев 等不等式.

设 $\{\xi_n\}$ 为随机变量序列，若对任意足标集 $i_1, \cdots, i_k$ 及实数 $x_{i_1}, \cdots, x_{i_k}$ 成立，

$$P(\omega: \xi_{i_1}<x_{i_1}, \cdots, \xi_{i_k}<x_{i_k})=\prod_{j=1}^{k} P(\xi_{i_j}<x_{i_j}),$$

则称 $\{\xi_n\}$ 相互独立.

独立性是一个重要概念，当 $\{\xi_n\}$ 相互独立时，可导出许多重要结论. 其中有些性质对比独立变量序列更广的一类序列也成立，这类序列就是下面要提到的鞅差序列.

§1.2 停时、鞅和鞅差序列

设 $t\geqslant 0$ 为离散或连续时间，$\{\mathscr{F}_t\}$ 为 σ-代数流，$\mathscr{F}_s\subset\mathscr{F}_t$，$\forall s<t$. 设对任意 $t\geqslant 0$，ξ_t 是随机变量，并且 ξ_t 对 $\mathscr{F}_t$ 可测，这时 $\{\xi_t, \mathscr{F}_t\}$ 叫适应过程. 还设 $\mathrm{E}|\xi_t|<\infty$，$t\geqslant 0$.

如果 $E(\xi_t|\mathscr{F}_s)=\xi_s, \forall t\geqslant s$，则$\{\xi_t, \mathscr{F}_t\}$叫鞅；如果 $E(\xi_t|\mathscr{F}_s)\leqslant\xi_s, \forall t\geqslant s$，则$\{\xi_t, \mathscr{F}_t\}$叫上鞅；如果 $E(\xi_t|\mathscr{F}_s)\geqslant\xi_s, \forall t\geqslant s$，则$\{\xi_t, \mathscr{F}_t\}$叫下鞅.

设 τ 为非负 $\mathscr{F}$-可测函数，如果对任一 $t\geqslant 0$，$\{\omega: \tau(\omega)\leqslant t\}\in\mathscr{F}_t$，则 τ 称为停时.

当 t 是连续时间时，下面设 $\mathscr{F}_t$ 右连续. 如果非负 $\mathscr{F}$-可测函数 τ，对任一 $t\geqslant 0$，$\{\omega: \tau(\omega)<t\}\in\mathscr{F}_t$，那么 τ 也是停时. 这是因为对任意 $\varepsilon>0$，$\{\omega: \tau(\omega)\leqslant t\}\in\mathscr{F}_{t+\varepsilon}$，所以$\{\omega: \tau(\omega)\leqslant t\}\in\mathscr{F}_{t^+}=\mathscr{F}_t$(因为 $\mathscr{F}_t$ 右连续).

引理1.2.1 设 τ_1, τ_2 为停时，则 $\tau_1\wedge\tau_2\triangleq\min(\tau_1, \tau_2)$，$\tau_1\vee\tau_2\triangleq\max(\tau_1, \tau_2)$及 $\tau_1+\tau_2$ 都是停时. 若 $\tau_1, \tau_2, \cdots$是停时，$\mathscr{F}_t$ 右连续，则 $\inf\limits_n\tau_n$、$\lim\limits_n\sup\tau_n$、$\lim\limits_n\inf\tau_n$ 也是停时.

对停时 τ，可定义到 τ 为止的事件的 σ-代数 $\mathscr{F}_\tau$:

$$\mathscr{F}_\tau=\{A: A\cap\{\tau\leqslant t\}\in\mathscr{F}_t, \forall t\geqslant 0, \forall A\in\mathscr{F}\}.$$

引理1.2.2 设 σ, τ 为停时，那么$\{\tau<\sigma\}$、$\{\tau>\sigma\}$、$\{\tau\leqslant\sigma\}$、$\{\tau\geqslant\sigma\}$及$\{\tau=\sigma\}$诸集同时属于 $\mathscr{F}_\sigma$ 及 $\mathscr{F}_\tau$. 设$(\xi_t, \mathscr{F}_t)$为适应过程，则 ξ_τ 对 $\mathscr{F}_\tau$ 可测.

引理1.2.3 设$(\xi_t, \mathscr{F}_t)$为实适应过程. ξ_t 和 $\mathscr{F}_t$ 都右连续. σ 为停时，那么 σ 后和 σ^+ 后首达开集 O 时间为停时.

证明 用 A^c 表集合 A 的余集.

σ 后首达 O 的时间为

$$\tau=\begin{cases}\inf\{t\geqslant\sigma: \xi_t\in O\},\\ 0, \quad \text{如果 } \xi_\sigma\in O.\end{cases}$$

σ^+ 后首达 O 的时间为

$$\mu=\begin{cases}\inf\{t>\sigma: \xi_t\in O\},\\ 0, \quad \text{如果 } \xi_\sigma\in O.\end{cases}$$

下面用 r 表示有理数. 由于

$$\{\omega: \tau\geqslant t\}=\{\omega: \xi_s\in O^c, \sigma\leqslant s<t\}=\bigcap_{r<t}\{\xi_r\in O^c, \sigma\leqslant r\}\in\mathscr{F}_t,$$

所以$\{\tau<t\}\in\mathscr{F}_t$，但 $\mathscr{F}_t$ 右连续，所以 τ 是停时. 同理可证 μ

也是停时.

引理1.2.4 设ξ_t对t连续(ξ_t的样本a.s.连续), D为R^1的闭集, 那么, $\tau_D \triangleq \inf(t \geqslant 0: \xi_t \in D)$相对于$\mathscr{F}_t^\xi = \sigma(\xi_s, s \leqslant t)$为停时.

引理1.2.5 设$(\xi_n, \mathscr{F}_n)$是离散时间的鞅(或上鞅或下鞅), τ为对$\mathscr{F}_n$的停时, 那么$(\xi_{n\wedge\tau}, \mathscr{F}_n)$也是鞅(或上鞅或下鞅). 对连续时间若$(\xi_t, \mathscr{F}_t)$, $t \geqslant 0$为具右连续轨线的鞅, τ为对$\mathscr{F}_t$的停时, 那么$(\xi_{t\wedge\tau}, \mathscr{F}_t)$也是鞅.

定理1.2.1 设$(\xi_t, \mathscr{F}_t)$为非负下鞅, $0 \leqslant t \leqslant T$, t为连续或离散时间, 且$E\xi_T^p < \infty$, $1 < p < \infty$, 则

$$E(\sup_{t \leqslant T} \xi_t)^p \leqslant \left(\frac{p}{p-1}\right)^p E\xi_T^p.$$

鞅的收敛定理在分析递推估计收敛性时是一个有力的工具.

定理1.2.2 设$(\xi_t, \mathscr{F}_t)$是具有右连续轨线的下鞅, 并且$\sup_t E\xi_t^+ < \infty$, 那么当$t \to \infty$时ξ_t a.s.收敛到有穷极限, $\xi_t \xrightarrow[t\to\infty]{} \xi < \infty$ a.s.并且$E\xi^+ < \infty$. 这里定理对离散时间也成立.

离散时间的适应过程$(x_n, \mathscr{F}_n)$若满足

$$E(x_n | \mathscr{F}_{n-1}) = 0, \ \forall n \geqslant 1,$$

则称为*鞅差序列*.

鞅差序列比相互独立随机变量序列要广, 它是一种常用的噪声模型.

定理1.2.3 设$\{x_n, \mathscr{F}_n\}$是鞅差序列. 那么当$n \to \infty$时, $\xi_n \triangleq \sum_{i=1}^{n} x_i$在

$$A \triangleq \left\{ \omega: \sum_{i=1}^{\infty} E(|x_i|^p | \mathscr{F}_{i-1}) < \infty \right\}, \quad 1 \leqslant p \leqslant 2$$

上收敛.

定理1.2.4 设$\{x_n, \mathscr{F}_n\}$是鞅差序列, $\{M_n, \mathscr{F}_n\}$是适应过程. 如果

$$\sup_n E(\|x_{n+1}\|^\alpha | \mathscr{F}_n) < \infty \quad \text{a.s.} \ \alpha \in (0, 2],$$

那么当 $n\to\infty$ 时

$$\sum_{i=0}^{n} M_i x_{i+1}=O(s_n(\alpha)\log^{\frac{1}{\alpha}+\eta}(s_n^{\alpha}(\alpha)+e)),\ \text{a. s.}\ \forall\eta>0,$$

$$s_n(\alpha)=\left(\sum_{i=0}^{n}\|M_i\|^{\alpha}\right)^{\frac{1}{\alpha}}$$

定理 1.2.5(Borel-Cantelli-Levy 引理) 设$\{B_i\}$是事件列，$B_i\in\mathscr{F}_i$, $\mathscr{F}_i\subset\mathscr{F}_j$, $i\leqslant j$, $i\geqslant 1$. 那么 $\sum_{i=1}^{\infty}I_{B_i}<\infty$ 的充分必要条件是 $\sum_{i=1}^{\infty}P(B_i|\mathscr{F}_{i-1})<\infty$, 换句话说,

$$\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}B_i=\left\{\sum_{i=1}^{\infty}P(B_i|\mathscr{F}_{i-1})=\infty\right\}.$$

推论 1.2.1(Borel-Cantelli 引理) 设$\{B_i\}$为事件列. 如果 $\sum_{i=1}^{\infty}P(B_i)<\infty$, 那么

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}B_i\right)=0.$$

如果 B_i 相互独立, 并且 $\sum_{i=1}^{\infty}P(B_i)=\infty$, 那么

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}B_i\right)=1.$$

在收敛性分析中, 另一个常用的事实是 Kronecker 引理.

引理 1.2.6(Kronecker 引理) 设 $b_i>0$, $i=1, 2, \cdots$, $b_i<b_j$, $\forall i\leqslant j$, $b_i\xrightarrow[i\to\infty]{}\infty$. 设$\{M_i\}$为一个矩阵列, 如果 $\sum_{i=1}^{\infty}\frac{1}{b_i}M_i<\infty$, 那么 $\frac{1}{b_n}\sum_{i=1}^{n}M_i\xrightarrow[n\to\infty]{}0$.

证明 记 $N_0=0$, $N_n=\sum_{i=1}^{n}\frac{1}{b_i}M_i$, $b_0=0$. 从引理条件知

$$N_n\xrightarrow[n\to\infty]{}N<\infty.$$

对任给 $\varepsilon>0$, 当 $i\geqslant n_0$ 时, $\|N_i-N\|<\varepsilon$. 所以

$$\left\|\frac{1}{b_n}\sum_{i=1}^{n}M_i\right\|=\left\|\frac{1}{b_n}\sum_{i=1}^{n}b_i(N_i-N_{i-1})\right\|$$

$$=\left\|N_n+\frac{1}{b_n}\sum_{i=2}^{n}(b_{i-1}-b_i)N_{i-1}\right\|$$

$$=\left\| N_n-\frac{b_n-b_1}{b_n}N+\frac{1}{b_n}\sum_{i=2}^{n}(b_{i-1}-b_i)(N_{i-1}-N)\right\|$$

$$\leqslant\|N_n-N\|+\frac{b_1}{b_n}\|N\|+\frac{1}{b_n}\sum_{i=2}^{n}(b_{i-1}-b_i)\|N_{i-1}-N\|$$

$$+\varepsilon\xrightarrow[\substack{n\to\infty\\ \varepsilon\to 0}]{}0.$$ □

§1.3 随机微分方程

设$(\omega_t,\mathscr{F}_t)$是l维轨线连续的鞅，如果$\omega_0=0, \mathrm{E}\omega_t=0, \mathrm{E}\|\omega_t\|^2<\infty$, $\forall t\geqslant 0$, 并且

$$\mathrm{E}[(\omega_t-\omega_s)(\omega_t-\omega_s)^\tau|\mathscr{F}_s]=(t-s)I,\ \forall t\geqslant s\geqslant 0,$$

那么$(\omega_t, \mathscr{F}_t)$叫 Wiener 过程。

Wiener 过程是独立增量过程，即$\omega_{t_1}-\omega_{t_2}$和$\omega_{t_3}-\omega_{t_4}$独立，$\forall t_1\geqslant t_2\geqslant t_3\geqslant t_4$, 并且$\omega_t-\omega_s$服从正态分布，

$$\limsup_{t\to\infty}\frac{\|\omega_t\|^2}{2lt\log\log t}=1, \qquad \text{a. s.}$$

设适应过程$(\xi_t, \mathscr{F}_t)$的轨线平方可积：

$$P\left(\int_0^T\|\xi_t\|^2dt<\infty\right)=1, \qquad T\leqslant\infty,$$

那么称$(\xi_t, \mathscr{F}_t)\in\mathscr{P}_T$. 对任一$(\xi_t, \mathscr{F}_t)\in\mathscr{P}_T$可定义积分

$$\int_0^T\xi_t d\omega_t.$$

为此先设ξ_t为简单过程：

$$\xi_t=\xi_0 I_{[0]}+\sum_{i=0}^{n-1}\xi_i I_{(t_i,t_{i+1}]},$$

$0=t_0<t_1<\cdots<t_n=T$, ξ_i对$\mathscr{F}_{t_i}$可测。

这时定义

$$\int_0^t\xi_s\mathrm{d}\omega_s\triangleq\sum_{i=0}^{m}\xi_i(\omega_{t_{i+1}}-\omega_{t_i})+\xi_{m+1}(\omega_t-\omega_{t_{m+1}}),$$

$$t_{m+1}<t\leqslant t_{m+2}.$$

现设$(\xi_t, \mathscr{F}_t)$的轨线平方积分的期望有穷：

$$\mathrm{E}\int_0^T\|\xi_t\|^2\mathrm{d}t<\infty.$$

这时存在$(\xi_t^n, \mathscr{F}_t)$, $\mathrm{E}\int_0^T\|\xi_t^n\|^2\mathrm{d}t<\infty$,并且

$$\mathrm{E}\int_0^T\|\xi_t-\xi_t^n\|^2\mathrm{d}t\xrightarrow[n\to\infty]{}0.$$

可证$\int_0^T\xi_t^n\mathrm{d}\omega_t$有均方极限,并定义

$$\int_0^T\xi_t\mathrm{d}\omega_t\triangleq\lim_{n\to\infty}\int_0^T\xi_t^n\mathrm{d}\omega_t,$$

$$\int_0^t\xi_s\mathrm{d}\omega_s\triangleq\int_0^T I_{[0,t]}\xi_s\mathrm{d}\omega_s.$$

对$(\xi_t, \mathscr{F}_t)\in\mathscr{P}_T$,存在$\xi_t^n$, $\mathrm{E}\int_0^T\|\xi_t^n\|^2\mathrm{d}t<\infty$,使

$$\int_0^T\|\xi_t-\xi_t^n\|^2\mathrm{d}t\xrightarrow[n\to\infty]{P}0,$$

而把$\int_0^T\xi_t\mathrm{d}\omega_t$定义为$\int_0^T\xi_t^n\mathrm{d}\omega_t$依概率收敛的极限:

$$\int_0^T\xi_t\mathrm{d}\omega_t\stackrel{P}{=}\lim_{n\to\infty}\int_0^T\xi_t^n\mathrm{d}\omega_t.$$

如果$\mathrm{E}\int_0^T\|\xi_t^i\|^2\mathrm{d}t<\infty$, $i=1,2$,那么鞅$\left(\int_0^t\xi_s^i\mathrm{d}\omega_s, \mathscr{F}_t\right)$是鞅,期望为0,并且

$$\mathrm{E}\left(\int_0^t\xi_s^1\mathrm{d}\omega_s\right)\left(\int_0^t\xi_s^2\mathrm{d}\omega_s\right)=\int_0^t\mathrm{E}\xi_s^1(\xi_s^2)^\tau\mathrm{d}s.$$

设$(a_t, \mathscr{F}_t)$, $(b_t, \mathscr{F}_t)$是适应过程, $(b_t, \mathscr{F}_t)\in\mathscr{P}_T$, a_s的轨线在$[0, T]$上可积. 设

$$\xi_t=\xi_0+\int_0^t a_s\mathrm{d}s+\int_s^t b_s\mathrm{d}\omega_s.$$

$(\xi_t, \mathscr{F}_t)$叫伊滕过程. 它可形式地写成微分形式:

$$\mathrm{d}\xi_t=a_t\mathrm{d}t+b_t\mathrm{d}\omega_t$$

不言而喻,如果$f(t, x)$是可微函数,如何求$\mathrm{d}f(t, \xi_t)$是十分重要的. 这就是伊滕公式.

定理1.3.1 设$\{a_t, \mathscr{F}_t\}$是l-维可测过程(即(t, ω)两个变量

的可测函数), $\|a_t\|^{\frac{1}{2}}\in\mathscr{P}_T$, $(B_t,\mathscr{F}_t)$是 $l\times m$ 阵可测过程, $\|B_t\|\in\mathscr{P}_T$, $\{\omega_t,\mathscr{F}_t\}$是 m 维 Wiener 过程,

$$\mathrm{d}\xi_t=a_t\mathrm{d}t+B_t\mathrm{d}\omega_t,$$

又记 $x=[x^1\cdots x^l]^\tau$, $f_t(t,x)=\dfrac{\partial f(t,x)}{\partial t}$,

$$f_x(t,x)\triangleq\begin{bmatrix}\dfrac{\partial f(t,x)}{\partial x^1}\\ \vdots\\ \dfrac{\partial f(t,x)}{\partial x^l}\end{bmatrix},\quad f_{xx}\triangleq\begin{bmatrix}\dfrac{\partial^2 f(t,x)}{\partial x^1\partial x^1}\cdots\dfrac{\partial^2 f(t,x)}{\partial x^l\partial x^1}\\ \vdots\qquad\qquad\vdots\\ \dfrac{\partial^2 f(t,x)}{\partial x^1\partial x^l}\cdots\dfrac{\partial^2 f(t,x)}{\partial x^l\partial x^l}\end{bmatrix}.$$

又设$f_t(t,x)$, $f_x(t,x)$及$f_{xx}(t,x)$为连续函数, 那么

$$\mathrm{d}f(t,\xi_t)=[f_t(t,\xi_t)+f_x^\tau(t,\xi_t)a_t+\frac{1}{2}\mathrm{tr}f_{xx}(t,\xi_t)B_tB_t^\tau]\mathrm{d}t$$
$$+f_x^\tau(t,\xi_t)B_t\mathrm{d}\omega_t.$$

设$(C_T,\mathscr{B}_T)$是定义在$[0,T]$上连续函数的可测空间, $\mathscr{B}_t=\sigma\{x;x_s,s\leqslant t\}$是 $\mathscr{B}_T$ 的子 σ-代数, 即定义在$[0,t]$上连续函数所生成的最小 σ-代数.

设 $a_t(x)$及 $b_t(x)$对任一 $t\in[0,T]$为 $\mathscr{B}_t$-可测. 如果适应过程 $\{\xi_t,\mathscr{F}_t\}$ 使 $\|a_t(\xi)\|^{\frac{1}{2}}\in\mathscr{P}_T$, $b_t(\xi)\in\mathscr{P}_T$, $\eta\in\mathscr{F}_0$, 并对任意 $t\in[0,T]$, 有

$$\xi_t=\eta+\int_0^t a_s(\xi)\mathrm{d}s+\int_0^t b_s(\xi)\mathrm{d}\omega_s,\quad \text{a.s.},$$

那么称$(\xi_t,\mathscr{F}_t)$为随机微分方程

$$\mathrm{d}\xi_t=a_t(\xi)\mathrm{d}t+b_t(\xi)\mathrm{d}\omega_t,\ \xi_0=\eta \tag{1.3.1}$$

的强解.

现在来看线性随机微分方程

$$\mathrm{d}\xi_t=A_t\xi_t\mathrm{d}t+b_t\mathrm{d}t+D_t\mathrm{d}\omega_t,$$

$(A_t,\mathscr{F}_t)$, $(b_t,\mathscr{F}_t)$, $(D_t,\mathscr{F}_t)$都是相应维数的适应过程, $\|b_t\|\in\mathscr{P}_T$, $\|D_t\|\in\mathscr{P}_T$, $\|A_t\|^{\frac{1}{2}}\in\mathscr{P}_T$. 设 $\Phi_{t,s}$为基本解阵:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t,s}=A_t\Phi_{t,s},\ \Phi_{s,s}=I,\ \forall t\geqslant s,\quad \text{a.s.}$$

那么上述线性随机微分方程有唯一强解

$$\xi_t=\Phi_{t,0}\xi_0+\int_0^t\Phi_{t,s}b_s\mathrm{d}s+\Phi_{t,0}\int_0^t\Phi_{s,0}^{-1}D_s\mathrm{d}\omega_s.$$

对一般的非线性随机微分方程,强解不一定存在.

定理1.3.2 如果$a_t(x)$, $b_t(x)$二元可测, 满足 Lipschitz 条件

$$\|a_t(x)-a_t(y)\|+\|b_t(x)-b_t(y)\|\leqslant c\|x-y\|,$$
$$\forall t\in[0,\ T],\ \forall x,\ y\in C_T, \quad (1.3.2)$$

c为常数, 并且$a(t,\ 0)$, $b(t,\ 0)$对$t\in[0,\ T]$有界, 那么方程(1.3.1)在$[0,\ T]$上存在唯一强解.

Lipschitz 条件(1.3.2)对$a_t(x)$和$b_t(x)$要求很强, 对x的增长速度不能高于线性. 下面的条件要弱得多, 称为局部 Lipschitz 条件:

对任意n及T,当$t\leqslant T$时有

$$(\|a_t(x)-a_t(y)\|+\|b_t(x)-b_t(y)\|)I_{[|x|<n,|y|<n]}\leqslant$$
$$C_{n,T}\|x-y\|, \quad \forall x,\ y\in C_T,$$

这里常数$C_{n,T}$可依赖于n及T.

定理1.3.3 设$a_t(x)$, $b_t(x)$二元可测, $a_t(x)$及$b_t(x)$满足局部 Lipschitz 条件,那么方程(1.3.1)存在唯一局部强解ξ_t, $0\leqslant t<\zeta$, ζ叫爆炸时间,ξ_t在$(\zeta-\varepsilon,\ \zeta)$上 a.s. 无界, $\forall\varepsilon>0$.

我们要用到下面较一般形式的 Bellman-Gronwall 不等式.

定理1.3.4 设可测函数$g_t\geqslant0$, $f_t\geqslant0$, $h_t\geqslant0$,并且

$$g_t\leqslant f_t+\int_{t_0}^t h_s g_s\mathrm{d}s<\infty,\ \forall t\geqslant t_0,$$

那么

$$g_t\leqslant f_t+\int_{t_0}^t h_s f_s e^{\int_s^t h_\lambda\mathrm{d}\lambda}\mathrm{d}s. \quad (1.3.3)$$

证明 $$g_t\leqslant f_t+\int_{t_0}^t h_s\left(f_s+\int_{t_0}^s h_\lambda g_\lambda\mathrm{d}\lambda\right)\mathrm{d}s$$

$$\leqslant f_t+\int_{t_0}^t h_s f_s\mathrm{d}s+\int_{t_0}^t h_s\int_{t_0}^s h_\lambda\left(f_\lambda+\int_{t_0}^\lambda h_\mu g_\mu\mathrm{d}\mu\right)\mathrm{d}\lambda\mathrm{d}s$$

继续迭代下去，对重积分项表达如下：

$$\int_{t_0}^{t} h_s \int_{t_0}^{s} h_\lambda f_\lambda \mathrm{d}\lambda\, \mathrm{d}s = \int_{t_0}^{t} h_\lambda f_\lambda \mathrm{d}\lambda \int_{\lambda}^{t} h_s \mathrm{d}s,$$

$$\int_{t_0}^{t} h_s \int_{t_0}^{s} h_\lambda \int_{t_0}^{\lambda} h_\mu f_\mu \mathrm{d}\mu \mathrm{d}\lambda\, \mathrm{d}s = \int_{t_0}^{t} h_s \int_{t_0}^{s} h_\mu f_\mu \int_{\mu}^{t} h_\lambda\, \mathrm{d}\lambda\, \mathrm{d}\mu\, \mathrm{d}s$$

$$= \int_{t_0}^{t} h_\mu f_\mu\, \mathrm{d}\mu \int_{\mu}^{t} h_s \int_{\mu}^{s} h_\lambda\, \mathrm{d}\lambda\, \mathrm{d}s = \int_{t_0}^{t} h_\mu f_\mu \frac{\left(\int_{\mu}^{t} h_s\, \mathrm{d}s\right)^2}{2} \mathrm{d}\mu.$$

用归纳法可验证

$$\int_{t_0}^{t} h_{s_1} \int_{t_0}^{s_1} h_{s_2} \cdots \int_{t_0}^{s_{n-1}} h_{s_n} f_{s_n} \mathrm{d}s_n \cdots \mathrm{d}s_1 = \int_{t_0}^{t} h_s f_s \frac{\left(\int_{s}^{t} h_\lambda\, \mathrm{d}\lambda\right)^{n-1}}{(n-1)!},$$

所以

$$g_t \leqslant f_t + \int_{t_0}^{t} h_s f_s \left(1 + \int_{s}^{t} h_\lambda \mathrm{d}\lambda + \frac{\left(\int_{s}^{t} h_\lambda\, \mathrm{d}\lambda\right)^2}{2!} + \cdots + \frac{\left(\int_{s}^{t} h_\lambda\, \mathrm{d}\lambda\right)^n}{n!}\right) \mathrm{d}s + \varphi_n(t).$$

注意到$\int_{t_0}^{t} h_s g_s \mathrm{d}s < \infty$，把 $n \to \infty$ 便得式(1.3.3)． □

推论 1.3.1 若 f_t 非降 $t \geqslant t_0$，那么从(1.3.3)知

$$g_t \leqslant f_t \left(1 + \int_{t_0}^{t} h_s e^{\int_s^t h_\lambda \mathrm{d}\lambda}\, \mathrm{d}s\right) = f_t e^{\int_{t_0}^t h_\lambda \mathrm{d}\lambda}.$$

对离散时间也有类似结果．

定理 1.3.5 设$\{g_k\}$，$\{f_k\}$及$\{h_k\}$为非负序列，且

$$g_k \leqslant f_k + \sum_{i=0}^{k-1} h_i g_i, \quad k \geqslant 0, \tag{1.3.4}$$

那么

$$g_k \leqslant f_k + \sum_{i=0}^{k-1} \prod_{j=i}^{k-1} (1 + h_j) f_i, \quad k \geqslant 0. \tag{1.3.5}$$

证明 记

$$Y_k = \sum_{i=0}^{k} h_i g_i, \quad Y_{-1} = 0,$$

那么

$$
\begin{aligned}
Y_k &= h_k g_k + \sum_{i=0}^{k-1} h_i g_i \leqslant h_k\left(f_k + \sum_{i=0}^{k-1} h_i g_i\right) + \sum_{i=0}^{k-1} h_i g_i \\
&= (1+h_k) Y_{k-1} + h_k f_k.
\end{aligned}
$$

所以
$$
Y_k \leqslant \sum_{i=0}^{k} \prod_{j=i+1}^{k} (1+h_j) h_i f_i \leqslant \sum_{i=0}^{k} \prod_{j=i}^{k} (1+h_j) f_i.
$$

这和式(1.3.4)一起，便得式(1.3.5)． □

§1.4　阵的伪逆与微分

这一节我们叙述将要用到的矩阵伪逆和矩阵微分．除对伪逆有简单证明外，其余未加证明，读者可参考文献[16, 45]．

设 $A \in C^{n\times m}$，这里 $C^{n\times m}$ 表复域上的 $(n\times m)$ 阵空间．矩阵的伪逆是用满足下面联立的矩阵代数方程（又称为 Penrose-Moore 方程）来定义的，这些方程是：

$$
\begin{cases}
AXA = A, & (1.4.1) \\
XAX = X, & (1.4.2) \\
AX = (AX)^*, & (1.4.3) \\
XA = (XA)^*. & (1.4.4)
\end{cases}
$$

这里"$*$"表示转置兼取复数共轭，X 是未知阵．

如果 $n=m$，并且阵 A 的行列式 $\det A \neq 0$，那么 $X = A^{-1}$ 是满足上述联立方程的唯一解．对一般的 A 阵，有下面的定理．

定理1.4.1　矩阵代数方程(1.4.1)～(1.4.4)的解存在且唯一，这里解记作 A^+，它是 $m\times n$ 阵，称为 A 阵的伪逆．

证明　设 A 阵的秩为 r，$r \leqslant \min(n, m)$，于是 A 阵的列张成 n 维空间中的一个秩为 r 的线性子空间 $L(A)$，用 $n\times r$ 阵 B 的列表示 $L(A)$ 的基．由于 A 阵的任一列必是 B 阵列的线性组合，所以存在一个 $r\times m$ 阵 C，使得：

$$
A = BC. \tag{1.4.5}
$$

因为 A 的秩为 r，所以 B 和 C 都是满秩阵，我们称这样的矩阵分解叫满秩分解．

很容易直接验证

$$A^{+}=C^{*}(CC^{*})^{-1}(B^{*}B)^{-1}B^{*} \tag{1.4.6}$$

满足方程(1.4.1)～(1.4.4).

为证唯一性，设 X 和 Y 都满足(1.4.1)～(1.4.4)，那么

$$\begin{aligned} X &= X(AX)^{*}=XX^{*}A^{*}=XX^{*}(AYA)^{*} \\ &= X(AX)^{*}(AY)^{*}=XAY=XAYAY \\ &= (XA)^{*}(YA)^{*}Y=(YAXA)^{*}Y=(YA)^{*}Y \\ &= YAY=Y. \end{aligned}$$

□

定理1.4.2 伪逆有以下性质：

1) $(A^{*})^{+}=(A^{+})^{*}$, $(A^{+})^{+}=A$,

2) $(A^{*}A)^{+}=A^{+}(A^{*})^{+}=A^{+}(A^{+})^{*}$,

3) $A^{+}=(A^{*}A)^{+}A^{*}=A^{*}(AA^{*})^{+}$,

4) $A^{+}AA^{*}=A^{*}AA^{+}=A^{*}$,

5) 设 U 和 V 分别是 $C^{n\times n}$ 和 $C^{m\times m}$ 中的酉阵，那么

$$(UAV)^{+}=V^{*}A^{+}U^{*}.$$

证明 1) 利用 A 的满秩分解式(1.4.5)知 $A^{*}=C^{*}B^{*}$ 是 A^{*} 的满秩分解，所以利用式(1.4.6)知

$$(A^{*})^{+}=B(B^{*}B)^{-1}(CC^{*})^{-1}C=(A^{+})^{*}.$$

由于(1.4.6)式正好就是 A^{+} 的满秩分解，于是易验证 $(A^{+})^{+}=A$.

2) 记 $\bar{B}=C^{*}$, $\tilde{C}=B^{*}BC$，那么 $A^{*}A=\bar{B}\tilde{C}$ 是满秩分解，所以

$$\begin{aligned} (A^{*}A)^{+} &= C^{*}B^{*}B(B^{*}BCC^{*}B^{*}B)^{-1}(CC^{*})^{-1}C \\ &= C^{*}(CC^{*})^{-1}(B^{*}B)^{-1}(CC^{*})^{-1}C \\ &= A^{+}(A^{+})^{*}=A^{+}(A^{*})^{+}. \end{aligned}$$

4) 可以利用 A 阵的满秩分解得到，而5) 可以直接以方程(1.4.1)～(1.4.4)验证. □

定理1.4.3 矩阵代数方程 $AX=D$ 相容的充要条件是 $AA^{+}D=D$. 当此条件成立时，方程 $AX=D$ 的通解为

$$X=A^{+}D+(I-A^{+}A)Y\ (Y\text{ 为任意维数相容阵}).$$

我们现在介绍矩阵微分公式中要用到的一些记号和运算，并

不加证明地给出与本书内容(第8章)有关或常用的矩阵微分公式.

用 $E_{ij}(m\times n)$ 或 E_{ij} 表示一个 $(m\times n)$ 阶矩阵, 它的第 (i, j) 元素为 1, 其余的元素全为 0. 设矩阵 A 是 $(m\times n)$ 阶, 记它的第 i 列矢量为 $A_{\cdot i}$, $i=1, \cdots, n$. 定义

$$\operatorname{vec} A=\begin{pmatrix} A_{\cdot 1} \\ A_{\cdot 2} \\ \vdots \\ A_{\cdot n} \end{pmatrix}. \tag{1.4.7}$$

我们称 vec 为矢量算符, 显然 $\operatorname{vec}A$ 是一个 mn 维矢量.

由上面的记号, 可以引出与 $\operatorname{vec}A$ 和 $\operatorname{vec}A^\tau$ 相联系的置换阵的定义, 这里"τ"表示转置.

矩阵 $A=[a_{ij}]$ 是 $(m\times n)$ 阶阵, 它可写为

$$A=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}E_{ij},$$

因此

$$A^\tau=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}E_{ij}^\tau,$$

$$\begin{aligned}\operatorname{vec} A^\tau&=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\operatorname{vec} E_{ij}^\tau\\ &=[\operatorname{vec} E_{11}^\tau \operatorname{vec} E_{21}^\tau\cdots\operatorname{vec} E_{m1}^\tau \operatorname{vec} E_{12}^\tau\cdots\operatorname{vec} E_{mn}^\tau]\begin{bmatrix} a_{11}\\ a_{21}\\ \vdots\\ a_{m1}\\ a_{12}\\ \vdots\\ a_{mn}\end{bmatrix}\\ &=[\operatorname{vec} E_{11}^\tau \operatorname{vec} E_{21}^\tau\cdots\operatorname{vec} E_{mn}^\tau]\operatorname{vec} A.\end{aligned}$$

于是我们记

$$U=[\operatorname{vec} E_{11}^\tau \operatorname{vec} E_{21}^\tau\cdots\operatorname{vec} E_{m1}^\tau \operatorname{vec} E_{12}^\tau\cdots\operatorname{vec} E_{mn}^\tau], \tag{1.4.8}$$

也就是

$$\text{vec}\, A^T = U\, \text{vec}\, A, \tag{1.4.9}$$

U 称为置换矩阵.

显然, U 是一个$(mn\times mn)$阶的方阵. 可以证明它是直交阵:

$$U^{-1} = U^T. \tag{1.4.10}$$

我们给出在矩阵微分中十分有用的矩阵运算——Kronecker积的定义及其性质.

设有$(m\times n)$阶矩阵 $A=[a_{ij}]$和$(r\times s)$阶矩阵 $B=[b_{ij}]$, 它们的 Kronecker 积表示成 $A\otimes B$, 定义为如下分块阵

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ & \cdots\cdots\cdots & & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}. \tag{1.4.11}$$

易见 $A\otimes B$ 是一个$(mr\times ns)$阶的阵, 它有 mn 个块, 第(i, j)块是一个$(r\times s)$阶的阵 $a_{ij}B$.

定理 1.4.4 Kronecker 积有如下的性质和运算法则:

1) 数乘交换律:

若 α 是一个数, 则

$$A\otimes(\alpha B)=\alpha(A\otimes B). \tag{1.4.12}$$

2) 对加法的分配律:

(a) $(A+B)\otimes C=A\otimes C+B\otimes C,$ (1.4.13)

(b) $A\otimes(B+C)=A\otimes B+A\otimes C.$ (1.4.14)

3) 结合律:

$$A\otimes(B\otimes C)=(A\otimes B)\otimes C. \tag{1.4.15}$$

4) 交换律: 设 A 和 B 分别为$(m\times n)$和$(r\times s)$阵, 则

$$A\otimes B=U_1(B\otimes A)U_2, \tag{1.4.16}$$

其中 U_1 和 U_2 分别为$(rm\times rm)$和$(sn\times sn)$阶的置换阵.

5) 混合积法则: 只要矩阵的维数使下列运算可行, 则有

$$(A\otimes B)(C\otimes D)=AC\otimes BD. \tag{1.4.17}$$

6) 积的逆阵法则: 只要 A, B 都是可逆方阵, 则

$$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}. \tag{1.4.18}$$

7）积的转置法则：

$$(A\otimes B)^T=A^T\otimes B^T. \tag{1.4.19}$$

8）积的特征值与特征矢量法则：设 $\{\lambda_i\}$ 和 $\{x_i\}$ 是阵 A 的特征值及对应的特征矢量，$\{\mu_j\}$ 和 $\{y_j\}$ 是阵 B 的特征值及对应的特征矢量. 则 $A\otimes B$ 的特征值为 $\{\lambda_i\mu_j\}$，对应的特征矢量为 $\{x_i\otimes y_j\}$.

9）行列式法则：设 A 和 B 分别为 $(m\times m)$ 和 $(n\times n)$ 阶方阵，则

$$\det(A\otimes B)=(\det A)^m(\det B)^n. \tag{1.4.20}$$

10）迹法则：

$$\mathrm{tr}(A\otimes B)=\mathrm{tr}\,A\cdot\mathrm{tr}\,B. \tag{1.4.21}$$

11）向量算符法则：

$$\mathrm{vec}(AYB)=(B^T\otimes A)\,\mathrm{vec}\,Y. \tag{1.4.22}$$

12）置换阵的 Kronecker 积表示：式(1.4.8)可重写成

$$U=\sum_{i=1}^{m}\sum_{j=1}^{n}E_{ij}\otimes E_{ij}^T. \tag{1.4.23}$$

下面，我们将给出矩阵微分的定义，然后给出本书将用到的矩阵微分的部分公式.

设 $Y=[y_{ij}]$ 是一个 $(p\times q)$ 阶阵，Y 关于数 x 的导数定义为 $(p\times q)$ 阶阵 $[\partial y_{ij}/\partial x]$，记为 $\dfrac{\partial Y}{\partial x}$.

又设 $X=[x_{rs}]$ 是一个 $(m\times n)$ 阶阵，阵 Y 关于阵 X 的导数记为 $\dfrac{\partial Y}{\partial X}$，定义为

$$\frac{\partial Y}{\partial X}=\begin{bmatrix}\dfrac{\partial Y}{\partial x_{11}} & \dfrac{\partial Y}{\partial x_{12}} & \cdots & \dfrac{\partial Y}{\partial x_{1n}}\\ \dfrac{\partial Y}{\partial x_{21}} & \dfrac{\partial Y}{\partial x_{22}} & \cdots & \dfrac{\partial Y}{\partial x_{2n}}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial Y}{\partial x_{m1}} & \dfrac{\partial Y}{\partial x_{m2}} & \cdots & \dfrac{\partial Y}{\partial x_{mn}}\end{bmatrix}=\sum_{r=1}^{m}\sum_{s=1}^{n}E_{rs}\otimes\frac{\partial Y}{\partial x_{rs}}, \tag{1.4.24}$$

这里的 E_{rs} 与 X 同阶，即$(m\times n)$阶，而 $\partial Y/\partial X$ 为$(mp\times nq)$阶阵. 由于矢量是矩阵的特例，因而读者很容易写出矩阵关于矢量导数的定义和表达式. 下面，我们给出矩阵关于矩阵求导的链式法则，见文献[45].

定理1.4.5 设矩阵 Z 是阵 Y 的矩阵函数，而阵 Y 又是阵 X 的矩阵函数，即

$$Z=Z(Y(X)),$$

其中 $X=[x_{ij}]$ 是$(m\times n)$阶阵，$Y=[y_{ij}]$ 是$(u\times v)$阶阵，$Z=[z_{ij}]$ 是$(p\times q)$阶阵. 则

$$\frac{\partial Z}{\partial X}=\left[\frac{\partial[\mathrm{vec}Y]^{\tau}}{\partial X}\otimes I_p\right]\left[I_n\otimes\frac{\partial Z}{\partial\,\mathrm{vec}\,Y}\right], \quad (1.4.25)$$

其中 I_p 和 I_n 分别为$(p\times p)$和$(n\times n)$阶的单位阵.

定理1.4.6 以下是矩阵求导的部分公式:

1) $$\frac{\partial\,\mathrm{vec}\,(A\times B)}{\partial\,\mathrm{vec}\,X}=B^{\tau}\otimes A, \quad (1.4.26)$$

2) $$\frac{\partial\,\mathrm{vec}\,(X^{\tau}AX)}{\partial\,\mathrm{vec}\,X}=U^{\tau}(AX\otimes I)+(I\otimes A^{\tau}X), \quad (1.4.27)$$

3) $$\frac{\partial\,\mathrm{vec}\,(AX^{-1}B)}{\partial\,\mathrm{vec}\,X}=-(X^{-1}B)\otimes(X^{-1})^{\tau}A^{\tau}, \quad (1.4.28)$$

4) $$\frac{\partial\log(\det X)}{\partial X}=(X^{-1})^{\tau}, \quad (1.4.29)$$

5) $$\frac{\partial(\det X)^{n}}{\partial X}=n(\det X)^{n}(X^{-1})^{\tau}, \quad (1.4.30)$$

6) $$\frac{\partial\,\mathrm{tr}(AX)}{\partial X}=A^{\tau}, \quad (1.4.31)$$

7) $$\frac{\partial\,\mathrm{tr}(X^{\tau}AXB)}{\partial X}=AXB+A^{\tau}XB^{\tau}, \quad (1.4.32)$$

8) $$\frac{\partial\,\mathrm{tr}(X^{n})}{\partial X}=nX^{n-1}, \quad (1.4.33)$$

9) $$\frac{\partial\,\mathrm{tr}(e^{X})}{\partial X}=e^{X}, \quad (1.4.34)$$

10) $$\frac{\partial\,\mathrm{tr}(AX^{-1}B)}{\partial X}=-(X^{-1}BAX^{-1})^{\tau}, \quad (1.4.35)$$

11) $\frac{\partial Y}{\partial X}=\sum\sum E_{rs}\otimes\frac{\partial Y}{\partial x_{rs}}$, ($E_{rs}$ 与 X 同阶) (1.4.36)

12) $\frac{\partial X}{\partial X}=\overline{U}+U-\sum E_{rr}\otimes E_{rr}$, 当 X 为对称阵, (1.4.37)

其中 $\overline{U}=\sum\sum E_{rs}\otimes E_{rs}$,

13) $\frac{\partial X}{\partial X}=\overline{U}$, 当 X 的元素互相独立, (1.4.38)

14) $\frac{\partial X^{\tau}}{\partial X}=U$, 当 X 的元素互相独立, (1.4.39)

15) $\frac{\partial(XY)}{\partial Z}=\frac{\partial X}{\partial Z}(I\otimes Y)+(I\otimes X)\frac{\partial Y}{\partial Z}$, (1.4.40)

16) $\frac{\partial X^{-1}}{\partial X}=-(I\otimes X^{-1})\overline{U}(I\otimes X^{-1})$, (1.4.41)

17) $\frac{\partial(X\otimes Y)}{\partial Z}=\frac{\partial X}{\partial Z}\otimes Y+[I\otimes U_1]\left[\frac{\partial Y}{\partial Z}\otimes X\right][I\otimes U_2]$. (1.4.42)

第 2 章

随机逼近算法的分析方法

本章将先介绍随机逼近所研究的问题，然后以 Robbins-Monro 算法为对象，介绍研究算法收敛性的各种常见方法(主要是鞅方法、常微分方程方法、Lyapunov 函数方法)以及各种方法所用的典型条件和相应的收敛性定理.

§2.1 随机逼近算法

对许多实际问题和理论问题的研究，常常归结为一个非线性方程组的求根问题，而且很久以来，人们发展了不少逐次逼近的数值解法，例如 Newton 切线法. 但是，在很多实际问题中，人们并不知道方程中非线性函数的形式，只可以对给定的自变量值，带随机误差地量测到函数值. 在这种情况下，怎样去求这个未知的非线性函数的零点，是一个从实践到理论都很重要的问题.

设未知函数为 $h(x)$, $x\in R^l$, $h(\cdot)\in R^l$，它的零点为 x^0，即

$$h(x^0)=0. \tag{2.1.1}$$

对 $h(\cdot)$ 可以在任意点 x 进行量测，但量测带有误差，若 x_n 为第 n 次量测时所取定的自变量值，则函数的观测值为

$$y_{n+1}=h(x_n)+\varepsilon_{n+1}, \tag{2.1.2}$$

$\{\varepsilon_n\}$ 是量测误差序列，可以依赖于 x_n. $h(\cdot)$ 称为回归函数.

用实际得到的数据序列 $\{x_n\}$ 和 $\{y_n\}$，去求回归函数的根 x^0，这就是随机逼近问题. 在随机系统的估计、预报、控制和优化中，问题常常归结为随机逼近. 下面用一个例子说明随机逼近的实际性.

假设发明了一种新药，必须确定它的最佳剂量. 设剂量为 x，人服药后的反应为 $f(x)$，而人们希望达到的最佳反应为 α，定义 $h(x)=f(x)-\alpha$. 显然，这里的 $f(\cdot)$ 和 $h(\cdot)$ 都只是一种形式上的写法，人们无法精确地知道药物剂量与药物反应之间对应关系 $h(\cdot)$，但可以带随机误差地量测它，得到

$$y_{n+1}=f(x_n)-\alpha+\varepsilon_{n+1}. \tag{2.1.3}$$

解决这一实际问题就要求助某种随机逼近算法.

1951 年，Robbins 和 Monro 首次提出并研究了一种随机逼近算法 [78]，他们取数列 $\{a_i\}$ 为增益系数:

$$a_i>0,\ \sum_{i=0}^{\infty}a_i=\infty,\quad \sum_{i=0}^{\infty}a_i^2<\infty. \tag{2.1.4}$$

对 x^0 的第 $n+1$ 次逼近为

$$x_{n+1}=x_n+a_n y_{n+1}, \tag{2.1.5}$$

其中 y_n 如式 (2.1.2) 所定义，这就是著名的 Robbins-Monro (RM) 算法. 当时，他们讨论了 ε_n 为相互独立的情况，并且 $l=1$，在 $h(\cdot)$ 严格单调时，他们证明了 x_n 对 x^0 的平方收敛性

$$E|x_n-x^0|^2\xrightarrow[n\to\infty]{}0.$$

虽然上面的结果是在较强条件下得到的，但这标志着人们找到了对理论和实际都很重要的这类问题的一种处理办法. 随机逼近问题的研究一直吸引着许多应用数学家、统计学家和系统与控制专家的关注，所涉及的领域有过程统计、系统辨识与参数估计、适应性控制、随机优化与决策和信号处理等.

在 RM 算法提出后的第二年，出现了 Kiefer-Wolfowitz (KW)算法 [59]，也就是求 $h(x)$ 极值的算法. 如果能直接量测 $h(\cdot)$ 的导数，那么问题就归结为上面的 RM 算法. 但有时只能量测 $h(\cdot)$ 本身，只好利用 $h(\cdot)$ 的量测值的差商去估计 $h(\cdot)$ 的导数值，这就是 KW 算法的基本思想. 因此，KW 算法与 RM 算法相比较，除了在量测点上有一些变化外，算法的基本形式及理论研究的基本思想都是一致的.

增益系数 $\{a_i\}$ 又称步长因子, 它所具有的性质 (2.1.4), 对随机逼近算法和其它某些随机递推算法都是必要的. 条件 $\sum_{i=0}^{\infty} a_i^2 < \infty$ 的实质是 a_i 必须趋于 0, 也就是每步的修正量应越来越小, 直到把量测误差的影响慢慢压制下去, 使 $a_i y_i$ 中的 $a_i \varepsilon_i \xrightarrow[i\to\infty]{} 0$. 但是 $\sum_{i=0}^{\infty} a_i = \infty$ 又说明, a_i 趋于 0 的速度不能太快. 因为如果 $\sum_{i=0}^{\infty} a_i < \infty$, 这时即使 $\varepsilon_i \equiv 0$, $h(\cdot)$ 一致有界, $\|h(\cdot)\| \leqslant c$, 那么

$$\sum_{i=0}^{\infty} \|x_{i+1} - x_i\| \leqslant \sum_{i=0}^{\infty} a_i \|h(x_i)\| \leqslant c \sum_{i=0}^{\infty} a_i.$$

这表明增量 $\|x_{i+1} - x_i\|$ 之和与初值 x_0 无关地一致有界, 因而当初值 x_0 与 x^0 相距很远时, x_i 不可能逼近 x^0, 即使有极限也不是 x^0. 这当然不是我们所希望的, 所以必须有 $\sum_{i=0}^{\infty} a_i = \infty$.

对任何一种递推算法, 收敛性是首要问题. 到 70 年代初, 在量测误差为鞅差序列时, 证明了各种意义下多维算法的收敛性, 如均方收敛和几乎处处收敛. 文献 [74] 比较全面地总结了 20 年的进展. 鞅差序列是一种不相关列, 由鞅差列 $\{\omega_n\}$ 生成的滑动平均 (MA) 列

$$e_n = A_0\omega_n + A_1\omega_{n-1} + \cdots + A_r\omega_{n-r} \tag{2.1.6}$$

是一种有限相关列, 而实际问题中的量测误差(噪声)列可能是无穷相关的, 如工程上常常用到的自回归滑动平均 (ARMA) 列

$$e_n + B_1 e_{n-1} + \cdots + B_q e_{n-q} = A_0\omega_n + A_1\omega_{n-1} + \cdots + A_r\omega_{n-r}. \tag{2.1.7}$$

直到 70 年代中, 对相关噪声下的随机逼近收敛性分析, 还没有合适的方法. 究其原因, 从研究方法上看, 人们主要用鞅收敛定理来证明算法的收敛性, 而当量测误差无穷相关时就破坏了应用鞅收敛定理的前提. 此外, 很多学者都认为随机收敛性问题自然应采用概率方法, 没有意识到可以借助随机收敛性以外的数学方法.

1977年，瑞典学者 Ljung L 在对随机递推算法作收敛性分析时，想到形如式(2.1.5)一类的算法的收敛性，与下面的常微分方程

$$\frac{\mathrm{d}x}{\mathrm{d}t}=h(x) \tag{2.1.8}$$

的稳定性有关. 粗略地说，记 $x(t_n)\triangleq x_n$，$t_{n+1}-t_n\triangleq a_n$，把算法(2.1.5)写成

$$\frac{x(t_{n+1})-x(t_n)}{t_{n+1}-t_n}=\frac{x_{n+1}-x_n}{a_n}=h(x(t_n))+\varepsilon_n, \tag{2.1.9}$$

由于随 a_n 趋于 0，ε_n 的作用将被压制下去，于是得到方程(2.1.8). Ljung 提出的利用常微分方程稳定性定理来证明算法收敛性的方法，简称为常微分方程(ODE)方法[65]，数学上较严密的叙述可见 Kushner 和 Clark 的专著[49]. 对随机逼近算法的进一步研究表明，证明算法收敛的本质不在于把数据内插成连续函数，也不在于寻找这些函数的"尾函数"所满足的微分方程，而在于存在和 $h(\cdot)$ 相应的 Lyapunov 函数 $v(\cdot)$. 从后面的章节可以看到，不必做数据的内插函数，可以证明 $v(x_n)$ 不可能无穷次穿越任一非零区间，从而可简捷地证明算法的大范围收敛性，同时，还可得到稳健性结果. 我们称这种方法为"Lyapunov 函数"方法.

§2.2 鞅方法

用鞅方法研究 RM 算法的收敛性，主要针对不相关量测噪声的情形. 这个方法的基本思路是：首先设法把要讨论的收敛性问题中的序列转化为某一个上鞅或鞅序列，然后证明这样的鞅序列满足鞅收敛定理的条件，获得鞅序列的收敛性，再转化成原问题要讨论的序列的收敛性. 为了说明这种方法，我们来证明下面的定理.

我们要用到下面假设：

A 2.2.1 步长因子列$\{a_n\}$满足

$$a_n>0,\ \sum_{i=1}^{\infty}a_i=\infty,\ \sum_{i=1}^{\infty}a_i^2<\infty. \tag{2.2.1}$$

A 2.2.2 存在具有有界连续二阶导数的 Lyapunov 函数 $v(x): R^l\to R$，满足以下条件

$$v(x)>0,\ \forall x\neq x^0,\ v(x^0)=0;\quad 当\|x\|\to\infty 时, v(x)\to\infty, \tag{2.2.2}$$

并且对任意 $\varepsilon>0$，有

$$\sup_{\|x-x^0\|>\varepsilon} v_x^{\mathrm{T}}(x)\cdot h(x)=-\beta_\varepsilon<0. \tag{2.2.3}$$

A 2.2.3 量测误差 $(\varepsilon_n, \mathscr{F}_n)$ 是鞅差序列（ε_n 可依赖于 $x_{n-1},\cdots,x_0$），满足

$$\mathrm{E}(\varepsilon_n|\mathscr{F}_{n-1})=0,\ \mathrm{E}\|\varepsilon_n\|^2<\infty. \tag{2.2.4}$$

A 2.2.4 回归函数 $h(x)$ 满足

$$\|h(x)\|^2+\mathrm{E}(\|\varepsilon_k\|^2|\mathscr{F}_{k-1})<c(1+v(x)),\ \forall k\geqslant 0. \tag{2.2.5}$$

在本书中常用到的 c, c_i, m 表常数，同样的 c 在不同地方可能不同.

定理 2.2.1 设 A 2.2.1～A 2.2.4 成立，则对任意给定的初值 x_0，由算法(2.1.5)所定义的 x_n 几乎处处收敛到 x^0.

在证明定理之前，先证 2 个引理.

设 G 是 R^l 中的 Borel 集，$\{\mathscr{F}_k\}$ 为非降 σ-代数族，定义 σ 为 $\{x_k\}$ 从 G 初出的时间(若 $x_0\in G$，定义 $\sigma=0$). 由于 $[\sigma\leqslant n]=[x_0\overline{\in}G]\cup[x_0\in G, x_1\overline{\in}G]\cup\cdots\cup[x_0\in G,\cdots,x_{n-1}\in G, x_n\overline{\in}G]$，所以 $[\sigma\leqslant n]\in\mathscr{F}_n$，即 σ 是停时.

引理 2.2.1 设 $v(x)\geqslant 0$, $v(x_k)$ 是非负上鞅，即

$$\mathrm{E}(v(x_k)|\mathscr{F}_{k-1})\leqslant v(x_{k-1}),$$

那么 $(v(x_{\sigma\wedge k}),\mathscr{F}_k)$ 也是非负上鞅，这里 $\sigma\wedge k=\min(\sigma, k)$.

证明 这是引理 1.2.4 的部分结论，它的证明如下：

$$v(x_{\sigma\wedge k})=v(x_\sigma)I_{[\sigma\leqslant k-1]}+v(x_k)I_{[\sigma>k-1]}.$$

由于 $$v(x_\sigma)I_{[\sigma\leqslant k-1]}=v(x_0)I_{[\sigma=0]}+\cdots+v(x_{k-1})I_{[\sigma=k-1]},$$

所以 $v(x_\sigma)I_{[\sigma\leqslant k-1]}$ 对 $\mathscr{F}_{k-1}$ 可测.

从下面一连串等式和不等式，即得引理结论：

$$\begin{aligned}\mathrm{E}(v(x_{\sigma\wedge k})|\mathscr{F}_{k-1}) &= v(x_\sigma)I_{[\sigma\leqslant k-1]}+I_{[\sigma>k-1]}\mathrm{E}(v(x_k)|\mathscr{F}_{k-1})\\ &\leqslant v(x_\sigma)I_{[\sigma\leqslant k-1]}+I_{[\sigma>k-1]}v(x_{k-1})\\ &= v(x_{\sigma\wedge(k-1)}).\end{aligned}$$ □

引理 2.2.2 设 $v(x)\geqslant 0$, $\forall x$, $(v(x_k),\mathscr{F}_k)$为非负上鞅，并且

$$\mathrm{E}(v(x_{k+1})|\mathscr{F}_k)\leqslant v(x_k)-\alpha_k I_{[\sigma>k]},$$

$$\alpha_k>0,\ \sum_{k=1}^{\infty}\alpha_k=\infty,\ \alpha_k \text{ 为实数},$$

那么 $P[\sigma<\infty]=1$, 即$\{x_k\}$在有限时间内跑出 G.

证明 设 $$v_{k+1}=v(x_{k+1})I_{[\sigma>k]}+\sum_{i=0}^{k}\alpha_i I_{[\sigma>i]},$$

那么 $$\mathrm{E}(v_{k+1}|\mathscr{F}_k)\leqslant (v(x_k)-\alpha_k)I_{[\sigma>k]}+\sum_{i=0}^{k}\alpha_i I_{[\sigma>i]}=v_k.$$

根据上鞅收敛定理1.2.2, 知 v_k a.s. 收敛到有穷极限,

所以 $$\sum_{i=0}^{\infty}\alpha_i I_{[\sigma>i]}<\infty \quad \text{a.s.}$$

即 $$P\left[\sum_{i=0}^{\infty}\alpha_i I_{[\sigma>i]}=\infty\right]=0.$$

但 $$[\sigma=\infty]\subset\left[\sum_{i=0}^{\infty}\alpha_i I_{[\sigma>i]}=\infty\right],$$

所以 $$P[\sigma=\infty]=0 \quad 或 \quad P[\sigma<\infty]=1.$$ □

现在证明定理 2.2.1.

由 A 2.2.3 和式(2.1.5)易见 x_n 是$\mathscr{F}_n$可测的. 再由 Taylor 展式

$$\begin{aligned}v(x_{n+1}) &= v(x_n)+a_n v_x^{\tau}(x_n)(h(x_n)+\varepsilon_{n+1})\\ &\quad+\frac{1}{2}a_n^2(h(x_n)+\varepsilon_{n+1})^{\tau}v_{xx}(\xi_n)(h(x_n)+\varepsilon_{n+1})\\ &\leqslant v(x_n)+a_n v_x^{\tau}(x_n)\varepsilon_{n+1}+c_1 a_n^2(\|h(x_n)\|^2\\ &\quad+\|\varepsilon_{n+1}\|^2)+a_n v_x^{\tau}(x_n)h(x_n).\end{aligned} \tag{2.2.6}$$

这里 ξ_n 的分量在 x_n 和 x_{n+1} 的相应分量之间.

因而由 $\mathrm{E}v(x_0)=v(x_0)<\infty$, 式(2.2.3)、式(2.2.4)、式(2.2.5)和 $\mathrm{E}v_x^{\tau}(x_n)\varepsilon_{n+1}=\mathrm{E}(v_x^{\tau}(x_n)\mathrm{E}(\varepsilon_{n+1}|\mathscr{F}_n))=0$ 可递推地得

到 $\mathrm{E}v(x_n)<\infty$, 以及

$$\mathrm{E}(v(x_{n+1})|\mathscr{F}_n)\leqslant v(x_n)+cc_1a_n^2(1+v(x_n))+a_nv_x^\tau(x_n)h(x_n). \tag{2.2.7}$$

下面, 构造一个非负上鞅 $(v_n, \mathscr{F}_n)$. 令

$$v_{n+1}\triangleq(1+v(x_{n+1}))\prod_{i=n+1}^{\infty}(1+cc_1a_i^2), \tag{2.2.8}$$

于是由式(2.2.7)得

$$\begin{aligned}\mathrm{E}(v_{n+1}|\mathscr{F}_n)&\leqslant[1+v(x_n)+cc_1a_n^2(1+v(x_n))\\&\quad+a_nv_x^\tau(x_n)h(x_n)]\prod_{i=n+1}^{\infty}(1+cc_1a_i^2)\\&=v_n+a_nv_x^\tau(x_n)h(x_n)\prod_{i=n+1}^{\infty}(1+cc_1a_i^2).\end{aligned} \tag{2.2.9}$$

用式(2.2.3)知 $\mathrm{E}(v_{n+1}|\mathscr{F}_n)\leqslant v_n$, 所以 v_n a.s. 收敛.

注意到 $\sum_{i=1}^{\infty}a_i^2<\infty$, 所以 $\prod_{i=k}^{\infty}(1+cc_1a_i^2)<\infty$, 因此从 v_n 收敛得知 $v(x_n)$ a.s. 收敛.

对任一 $\varepsilon>0$, 用 σ_ε^0 表示 $\{x_k\}$ 初出 $G_\varepsilon=\{x: \|x-x^0\|>\varepsilon\}$ 的时间, 用 σ_ε^i 表示在 σ_ε^{i-1} 后初出 G_ε 的时间: $\sigma_\varepsilon^i=\min\{t: t>\sigma_\varepsilon^{i-1}, x_t\in G_\varepsilon^c\}$, 那么对任意 i 用式 (2.2.3), 从式 (2.2.9) 还可得:

$$\begin{aligned}\mathrm{E}(v_{n+1}|\mathscr{F}_n)&\leqslant v_n+I_{[\sigma_\varepsilon^i>n]}a_nv_x^\tau(x_n)h(x_n)\\&\leqslant v_n-\beta_\varepsilon a_nI_{[\sigma_\varepsilon^i>n]}.\end{aligned}$$

用引理 2.2.2 知 $P[\sigma_\varepsilon^i<\infty]=1$, 即 x_k 一定进入 G_ε^c. 由于 i 任意, 所以存在子列 $\{x_{k_i}\}$, 当 $i\to\infty$ 时, $k_i\to\infty$, 使 $\|x_{k_i}-x^0\|\leqslant\varepsilon$, 又由于 ε 任意, 所以存在子列仍用 $\{x_{k_i}\}$ 表示, 使 $\|x_{k_i}-x^0\|\xrightarrow[i\to\infty]{}0$. 但已证 $v(x_k)$a. s. 收敛, 所以必须有 $v(x_k)\xrightarrow[k\to\infty]{}0$ a.s. 再用式(2.2.2), 由此知 $x_k\xrightarrow[k\to\infty]{}x^0$ a. s.. □

鞅方法通常要求所讨论的随机序列为鞅差序列, 如果这个序列为相关序列, 特别是无穷相关序列, 则很难用鞅方法去处理. 参考文献[74]是70年代以前用鞅方法研究随机逼近的总结.

§2.3 常微分方程方法

鞅方法对量测噪声要求较严，而常微分方程方法[65,49]，利用常微分方程稳定性定理来证明算法的收敛性，可处理相当广的一类噪声. 下面就是用常微分方程方法来证明随机逼近算法的大范围强收敛性.

常微分方程方法要用到分析数学中两个熟知的事实，叙述如下：

定理2.3.1 (Arzelà-Ascoli)[93] 设以$\lambda\in\Lambda$为足标的一族连续函数$\{f_\lambda(t)\}$，$t\in[0,\infty)$等度连续(即对任意$t\in[0,\infty)$及$\varepsilon>0$，存在$\delta>0$，只要$|t-s|<\delta$，则$|f_\lambda(t)-f_\lambda(s)|\leqslant\varepsilon$，$\forall\lambda\in\Lambda$)，并且一致有界，则一定存在连续函数$f(t)$及$\{f_\lambda(t)\}$的子序列$\{f_{\lambda_k}(t)\}$，对任一$t$的有穷区间，一致地有

$$|f_{\lambda_k}(t)-f(t)|\xrightarrow[k\to\infty]{}0.$$

考虑l维常微分方程

$$\dot{x}_t=h(x_t),\ t\geqslant 0;\ h(x^0)=0. \tag{2.3.1}$$

定理2.3.2 如果存在连续可微函数$v(x)>0$，$\forall x\neq x^0$；$v(x^0)=0$，当$\|x\|\to\infty$时，$v(x)\to\infty$，且对$\forall x\neq x^0$，

$$v_x^\tau(x)h(x)<0, \tag{2.3.2}$$

那么从任意初值开始，当$t\to\infty$时方程(2.3.1)的轨线必趋于x^0，也就是说，x^0是方程(2.3.1)的大范围渐近稳定解.

这个定理的证明见文献[48]的推论5.3，也可参阅文献[47, 64].

下面引进一组条件A 2.3：

A 2.3.1 $a_n>0,\ a_n\xrightarrow[n\to\infty]{}0,\ \sum_{i=1}^{\infty}a_i=\infty.$ (2.3.3)

A 2.3.2 存在二次连续可微的Lyapunov函数$v(x)$，满足以下条件

$$v(x)>0,\ \forall x\neq x^0,\ v(x^0)=0;\ \text{当}\ \|x\|\to\infty\ \text{时},\ v(x)\to\infty, \tag{2.3.4}$$

并且

$$v_x^{\mathrm{T}}(x)h(x)<0,\ \forall x\neq x^0. \tag{2.3.5}$$

A 2.3.3 量测误差序列 $\{\varepsilon_n\}$ 满足

$$\varepsilon_n=e_n+\nu_n,\ \text{其中}\ \sum_{n=1}^{\infty}a_ne_{n+1}\text{收敛},\ \nu_n\xrightarrow[n\to\infty]{}0,\quad \text{a.s.}. \tag{2.3.6}$$

注 2.3.1 显然, A 2.3.1 弱于 A 2.2.1, A 2.3.2 去掉了 A 2.2.2 中 $v_{xx}(\cdot)$ 的有界性. A 2.2.4 在这里不要求, 这因为 ODE 方法要预先假设算法 $\{x_n\}$ 的一致有界性, 也就是说 A 2.2.4 与假设算法一致有界在保证算法收敛性方面有相互替代的作用. 这里要强调 A 2.3.3 比 A 2.2.3 弱得多,满足 A 2.3.3 的$\{\varepsilon_n\}$不仅包含 A 2.2.3 所考虑的鞅差序列, 而且包含了很大一类无穷相关的随机序列, 甚至还可以通过 A 2.3.3 把 § 2.1 引言中提到的求回归函数极值的 KW 算法化为 RM 算法的特例来研究它的收敛性.

下面, 举例来说明式(2.3.6) 中的 $\{e_n\}$ 可以是一个无穷相关的随机序列.

例 2.3.1 设 $\{e_n\}$ 是一个由鞅差序列 $(\omega_n,\ \mathscr{F}_n)$ 所构成的无穷项滑动平均序列, 即

$$e_n=\omega_n+B_1\omega_{n-1}+\cdots+B_k\omega_{n-k}+\cdots,\ B_0=I, \tag{2.3.7}$$

其中

$$\mathrm{E}(\omega_{n+1}|\mathscr{F}_n)=0,\ \mathrm{E}(\|\omega_{n+1}\|^2|\mathscr{F}_n)\leqslant c_1,$$

$$\|B_n\|\leqslant c_2/n^r,\ \ \forall r>\frac{3}{2}.$$

还设 A 2.3.1 中的 $\{a_n\}$ 满足 $\sum\limits_{n=0}^{\infty}a_n^2<\infty$. 由这些假设, 容易验证, 对 $\forall n$, $\left(\sum\limits_{i=n}^{m}a_i\omega_{i-i_0},\ \mathscr{F}_{m-i_0}\right)$是一个平方可积鞅, 所以$\left(\left\|\sum\limits_{i=n}^{m}a_i\omega_{i-i_0}\right\|,\ \mathscr{F}_{n-i_0}\right)$是非负下鞅.

记 $\mu\triangleq\left(r+\frac{1}{2}\right)\Big/2>1$, $s=\sum\limits_{k=0}^{\infty}\frac{1}{k^{\mu}}$, 其中 $\frac{1}{0^{\mu}}\triangleq 0$. 于是在定理

1.2.1中取 $p=2$,便知对任意 $\varepsilon>0$,

$$P\left\{\sup_{m\geq n}\left\|\sum_{i=n}^{m}a_i e_{i+1}\right\|\geq\varepsilon\right\}=P\left\{\sup_{m\geq n}\left\|\sum_{k=0}^{\infty}B_k\sum_{i=n}^{m}a_i\omega_{i+1-k}\right\|\right.$$

$$\left.\geq\frac{\varepsilon}{s}\sum_{k=1}^{\infty}\frac{1}{k^{\mu}}\right\}\leq\sum_{k=0}^{\infty}P\left\{\sup_{m\geq n}\left\|B_k\sum_{i=n}^{m}a_i\omega_{i+1-k}\right\|\geq\frac{\varepsilon}{sk^{\mu}}\right\}$$

$$\leq c_3\sum_{k=0}^{\infty}\frac{1}{k^{2(r-\mu)}}\sum_{i=n}^{\infty}a_i^2E\|\omega_{i+1-k}\|^2\xrightarrow[n\to\infty]{}0, \tag{2.3.8}$$

所以存在子列 n_k, 使 $\sum_{k=1}^{\infty}P\left\{\sup_{m\geq n_k}\left\|\sum_{i=n}^{m}a_i e_{i+1}\right\|\geq\varepsilon\right\}<\infty$. 利用 Borel-Cantelli 引理(推论1.2.1),知 $\{e_n\}$ 满足条件 A2.3.3. 注意到一个稳定的自回归滑动平均 (ARMA) 序列可以写成如式(2.3.7), 且 $\{B_n\}$ 具有负指数衰减速度的序列. 因此, 稳定的 ARMA 序列也就成为式(2.3.7)所表示序列的特例. 在文献[49]中§2.2还例举了一些满足 A2.3.3 的 $\{e_n\}$ 的例子.

现在我们给出用常微分方程方法证明的 RM 算法大范围收敛性的定理.

定理2.3.3 设 $h(\cdot)$ 是值域和定义域都是 R^l 的连续函数, 并且 A2.3 成立, 还设算法 (2.1.5) 所给出 $\{x_n\}$ a.s. 一致有界, 则对任意给定的初值 x_0,

$$\lim_{n\to\infty}x_n=x^0 \quad \text{a.s.},$$

证明 将证明过程分成4步.

第1步 把 $\{x_n\}$ 通过线性插值变成连续函数, 然后逐次向左平移单位长度, 产生一族连续函数.

令

$$t_n=\sum_{i=0}^{n-1}a_i,\ t_0=0,\ m(t)=\max\{n: t_n\leq t\}. \tag{2.3.9}$$

定义连续函数 x_t^0,它是内插长度为 $\{a_n\}$ 的 x_n 的线性内插, 即

$$x_{t_n}^0\triangleq x_n,\ x_t^0=\frac{t_{n+1}-t}{a_n}x_n+\frac{t-t_n}{a_n}x_{n+1},\ t\in[t_n,\ t_{n+1}]. \tag{2.3.10}$$

类似地, 若记

$$q_n \triangleq \sum_{i=0}^{n-1} a_i e_{i+1}, \tag{2.3.11}$$

则 q_t^0 为内插长度为 $\{a_n\}$ 的 q_n 的线性内插.

再定义一连续函数族$\{x_n(t)\}$为

$$x_n(t) \triangleq x_{t+n}^0. \tag{2.3.12}$$

第 2 步　证明$\{x_n(t)\}$满足 Arzelà-Ascoli 定理的条件，得到一极限函数 $x(t)$.

设

$$\bar{x}_t = x_n,\ \bar{\nu}_t = \nu_n,\ t \in [t_n,\ t_{n+}\ ). \tag{2.3.13}$$

显然，由式(2.3.9)～式(2.3.12)知:

$$x_n(t) = x_0 + \int_0^{t+n} (h(\bar{x}_s) + \bar{\nu}_s)\mathrm{d}s + q_{t+n}^0, \tag{2.3.14}$$

且由定理假设，对任固定 ω，$\{x_n(t)\}$ 是一致有界的. 现证$\{x_n(t)\}$还是等度连续的.

对任 $\Delta>0$，我们有

$$\|x_n(t+\Delta) - x_n(t)\| \leqslant \left\| \int_{t+n}^{t+n+\Delta} (h(\bar{x}_s) + \bar{\nu}_s)ds \right\| + \|q_{t+n+\Delta}^0 - q_{t+n}^0\| \tag{2.3.15}$$

由 A2.3.3，q_t^0 随 $t\to\infty$ 趋于有穷极限，所以式(2.3.15)的右边后一项随 $\Delta\to 0$ 对 n 一致地趋于 0. 同时由 A2.3.3 和 $h(\bar{x}_t)$ 在 $t\in[0,\infty)$的有界性，式(2.3.15)的右边第 1 项也随 $\Delta\to 0$ 对 n 一致地趋于 0. 对 $\Delta<0$ 成立类似的推理，所以$\{x_n(t)\}$ 等度连续. 于是由 Arzelà-Ascoli 定理，$\{x_n(t)\}$ 中可以抽出一子列$\{x_{n_k}(t)\}$一致地在任意有限区间上收敛于一个连续函数 $x(t)$.

第 3 步　证明 $x(t)$满足方程

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h(x(t)) \tag{2.3.16}$$

从式(2.3.15)，A2.3.3 和 q_t^0 的收敛性知:

$$\frac{x(t+\Delta) - x(t)}{\Delta} = \frac{1}{\Delta} \lim_{k\to\infty} \left[\int_{t_{m(t+n_k)}}^{t_{m(t+n_k+\Delta)}} h(\bar{x}_s)\mathrm{d}s + \int_{t_{m(t+n_k+\Delta)}}^{t+n_k+\Delta} h(\bar{x}_s)\mathrm{d}s - \int_{t_{m(t+n_k)}}^{t+n_k} h(\bar{x}_s)\mathrm{d}s \right]$$

$$=\frac{1}{\Delta}\lim_{k\to\infty}\int_{t_{m(t+n_k)}}^{t_{m(t+n_k+\Delta)}}h(\bar{x}_s)\mathrm{d}s. \tag{2.3.17}$$

这最后一个等式成立是因为由式(2.3.9)知

$$t-t_{m(t)}<a_{m(t)}\xrightarrow[t\to\infty]{}0.$$

令

$$\delta_{ki}=\sum_{j=m(t+n_k)}^{m(t+n_k)+i-1}a_j,\ \forall i\in\{1,\ 2,\ \cdots,\ m(t+n_k+\Delta)-m(t+n_k)\}, \tag{2.3.18}$$

显然

$$\delta_{ki}\leqslant\Delta+a_{m(t+n_k)},$$

$$\begin{aligned}t+n_k-a_{m(t+n_k)}+\delta_{ki}&<t_{m(t+n_k)}+\delta_{ki}\\&=t_{m(t+n_k)+i}\leqslant t+n_k+\delta_{ki},\end{aligned} \tag{2.3.19}$$

于是存在 $\delta'_{ki}\in[0,\ \Delta+a_m(t+n_k)]$，对任 $i\in\{1,\ 2,\ \cdots,\ m(t+n_k+\Delta)-m(t+n_k)\}$，有

$$t_{m(t+n_k)+i}=n_k+t+\delta'_{ki}. \tag{2.3.20}$$

由式(2.3.17)、(2.3.20)可得

$$\begin{aligned}&\frac{x(t+\Delta)-x(t)}{\Delta}\\&=\frac{1}{\Delta}\Big[\lim_{k\to\infty}\sum_{i=m(t+n_k)}^{m(t+n_k+\Delta)-1}a_i[h(x_i)-h(x(t))]+\Delta h(x(t))\Big]\\&=h(x(t))+\frac{1}{\Delta}\lim_{k\to\infty}\sum_{i=0}^{m(t+n_k+\Delta)-m(t+n_k)-1}a_{m(t+n_k)+i}[h(x_{n_k}(t+\delta'_{ki}))\\&\quad-h(x_{n_k}(t))+h(x_{n_k}(t))-h(x(t))].\end{aligned} \tag{2.3.21}$$

再由式(2.3.19)和 $x_{n_k}(t)$ 的一致收敛性，上式右边随 $\Delta\to0$ 而趋于 $h(x(t))$，而且对 $\Delta<0$ 的情况也成立，这就证明了式(2.3.16)。

第4步 从 $x(t)\xrightarrow[t\to\infty]{}x^0$ 到 $x_n\xrightarrow[n\to\infty]{}x^0$.

由 A2.3.2 及定理 2.3.2 知 x^0 为 $x(t)$ 的大范围渐近稳定解。于是有

$$x(t)\xrightarrow[t\to\infty]{}x^0, \tag{2.3.22}$$

然后我们将充分利用式(2.3.22)完成定理的证明。

下面，用反证法证明 $\{x_n\}$ 的任一子序列都必须收敛于 x^0.

若存在一子列 $x_{n_k} \xrightarrow[k\to\infty]{} \bar{x} \neq x^0$，由式(2.3.22)，则对固定 ω，存在 t_ω，对任 $t \geqslant t_\omega$，有

$$\|x(t)-x^0\| < \|\bar{x}-x^0\|/4, \tag{2.3.23}$$

而且当 $k \geqslant k_1$，

$$\|x_{n_k}-\bar{x}\| < \|\bar{x}-x^0\|/4. \tag{2.3.24}$$

由式(2.3.10)和式(2.3.12)，我们重写

$$x_{n_k} = x^0_{t_{n_k}} = x^0_{[t_{n_k}-t_\omega]+t_\omega+\delta_k} = x_{[t_{n_k}-t_\omega]}(t_\omega+\delta_k), \tag{2.3.25}$$

其中 $[t_{n_k}-t_\omega]$ 表示 $t_{n_k}-t_\omega$ 的整数部分，因而 $0 \leqslant \delta_k < 1$. 不妨记 $[t_{n_k}-t]$ 为 m_k，则 $\{x_{m_k}(t)\}$ 也是一致有界和等度连续的，所以也可抽出一子列，不妨仍记为 $x_{m_k}(t)$，对任何有限区间一致地有

$$x_{m_k}(t) \xrightarrow[k\to\infty]{} x(t). \tag{2.3.26}$$

于是存在 k_2，当 $k \geqslant k_2$ 时有

$$\|x_{m_k}(t)-x(t)\| < \|\bar{x}-x^0\|/4, \ \forall t \in [t_\omega, \ t_\omega+1]. \tag{2.3.27}$$

从式(2.3.23)和式(2.3.27)又知，当 $k \geqslant k_2$ 时，

$$\|x_{m_k}(t)-x^0\| < \|\bar{x}-x^0\|/2, \ \forall t \in [t_\omega, \ t_\omega+1]. \tag{2.3.28}$$

这样，由式(2.3.25)和式(2.3.28)知，存在 k_3，使当 $k_3 \geqslant k_1$ 时，

$$\|x_{n_{k_3}}-x^0\| < \|\bar{x}-x^0\|/2. \tag{2.3.29}$$

注意到式(2.3.24)，由上式可得出

$$\|\bar{x}-x^0\| \leqslant \|\bar{x}-x_{n_{k_3}}\| + \|x_{n_{k_3}}-x^0\| < \left(\frac{1}{2}+\frac{1}{4}\right)\|\bar{x}-x^0\|$$
$$= \frac{3}{4}\|\bar{x}-x^0\|,$$

而这不可能，因此 $\{x_n\}$ 的任一子列都收敛于 x^0. 定理得证. □

在定理证明的第 2 步中，$\{x_n\}$的几乎处处一致有界的假设是不可少的，否则，就无法应用 Arzelà-Ascoli 定理. 因此 ODE 方法虽然放宽了对量测噪声的限制，但额外要求$\{x_n\}$有界是不尽人意的.

文献[18, 19]把 ODE 方法和鞅方法相结合，不再事先要求 $\{x_n\}$ 有界，对较一般的回归函数和不相关噪声，成功地进行了收敛性分析．ODE 方法要求 A 2.3.2 中的 Lyapunov 函数 $v(\cdot)$ 要满足常微分方程大范围渐近稳定性定理要求，这个条件还可能减弱，详见下面一节．

§2.4 Lyapunov 函数方法

上一节中，在算法一致有界的前提下，利用 A2.3.2 作为常微分方程大范围渐近稳定的充分条件，证明了算法的大范围收敛性．正如我们在 §2.3 末尾所说，ODE 方法虽然扩大了噪声的范围，但要事先假定算法有界，在下一章中，我们将克服这个不足．另一方面，ODE 方法比较繁琐，要求回归函数是连续的，并且不能在 A2.3.2 不严格满足时使用，不易用于算法稳健性分析．这一节，我们不用 ODE 方法，而直接用 Lyapunov 函数的性质．在算法有界性的假设下，来证明算法的收敛性．实际上，我们将证明比定理 2.3.3 更强一些的结果，为此引入条件 A2.4．

下面引入欧氏空间中两个集合之间的距离

$$d(S_1,\ S_2)=\inf\{\|x-y\|:\forall x\in S_1,\ y\in S_2\}. \tag{2.4.1}$$

A2.4.0 $h(\cdot)$ 为 $R^l\to R^l$ 的 Borel 可测函数，在有界集上有界，$x\in J$ 时，$h(x)=0$．

A2.4.1 和 A2.3.1 相同．

A2.4.2 a) 存在二次连续可微函数 $v(x)$ (不一定非负)：$R^l\to R$，使

对任意 $\Delta>\delta>0$，有

$$\sup_{\Delta\geq d(x,J)>\delta} h^\tau(x)v_x(x)<0. \tag{2.4.2}$$

b) $v(J)=\text{const}$；或者对任一 x，只要 $d(x,\ J)>0$，必有 $d(v(x),\ v(J))>0$，这里 $v(J)=\{y: y=v(x),\ \forall x\in J\}$．

记

$$m(k,\ T)=\max\left\{m:\ \sum_{i=k}^{m}a_i\leqslant T\right\}. \tag{2.4.3}$$

A2.4.3 当 $\{x_{n_k}\}$ 收敛时，则

$$\lim_{T\to0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{i=n_k}^{m(n_k,T_k)}a_i\varepsilon_{i+1}\right\|=0,\ \forall T_k\in[0,\ T]. \tag{2.4.4}$$

注 2.4.1 当 $h(\cdot)$连续时，显然式(2.4.2)等价于$h^\tau(x)v_x(x)<0,\ \forall x\overline{\in}J$. 当式(2.4.4)成立时，$a_{n_k}\varepsilon_{n_k+1}\xrightarrow[k\to\infty]{}0$.

注 2.4.2 和上一节条件相比，首先现在不要求 $h(\cdot)$连续，并且 $h(\cdot)$ 的零点不一定是单点，它可以是一个集合 J. 当 $h(\cdot)$ 连续，$J=x^0$ 时，显然式(2.4.2)等价于式(2.3.5)，并且 A2.4.2b) 一定成立. 我们也不要求当 $\|x\|\to\infty$ 时，$v(x)\to\infty$. 此外，当 A2.3.3 成立时，A2.4.3 一定成立.

如果 $n_k<m_k$, $v(x_{n_k})\leqslant\delta_1$, $v(x_{m_k})\geqslant\delta_2$, $\delta_1<v(x_i)<\delta_2$, $\forall i: n_k<i<m_k$, 那么称 $v(x_{n_k}),\ \cdots,\ v(x_{m_k})$穿越区间 $[\delta_1,\ \delta_2]$.

引理 2.4.1 设 $\{x_n\}$ 为随机矢量序列，$x_n\in R^l$. 固定样本 ω. 如果对 $\{x_n\}$ 的任一收敛子列 $\{x_{l_k}\}$，可找到 $c_1>0$, $T_1>0$ 及 k_T 使

$$\|x_m-x_{l_k}\|\leqslant c_1T,\ \forall T\in[0,\ T_1],\ \forall m: l_k\leqslant m\leqslant m(l_k,\ T),$$
$$\forall k\geqslant k_T, \tag{2.4.5}$$

并且 x_m, $l_k\leqslant m\leqslant m(l_k,\ T)$，由式(2.1.2)及式(2.1.5)定义. 那么在 A2.4 条件下(但不要求 A2.4.2b))，对任一区间 $[\delta_1,\ \delta_2]$, $\delta_1\leqslant\delta_2$, 只要 $d([\delta_1,\ \delta_2],\ v(J))>0$, 对这个固定的 ω,

i) 当 $\delta_1<\delta_2$ 时，"$v(x_n)$无穷次穿越 $[\delta_1,\ \delta_2]$" 及 "上面定义的穿越子列的起始点 $\|x_{n_k}\|$ 一致有界" 两者不可能同时成立.

ii) 当 $\delta_1=\delta_2$ 时，"$\{x_{n_k}\}$有界" 和 "$v(x_n)$收敛到 δ_1" 两者不可能同时成立.

证明 在证明引理之前，我们首先注意到，当 $\{x_n\}$ 有界时，那么在 A2.4.0 及 A2.4.3 条件下，式(2.4.5)一定成立.

i) 设存在 $[\delta_1,\ \delta_2]$, $d([\delta_1,\ \delta_2],\ v(J))>0$, $v(x_n)$ 无穷次穿越 $[\delta_1,\ \delta_2]$，即 $v(x_{n_k})\leqslant\delta_1$, $v(x_{m_k})\geqslant\delta_2$, $\delta_1<v(x_i)<\delta_2$, $\forall i: n_k<$

$i < m_k$，并且 $\|x_{n_k}\| \leqslant c$，$\forall k \geqslant 1$.

从 $\|x_{n_k}\| \leqslant c$，A2.4.0 及注 2.4.1 知，$x_{n_k+1} - x_{n_k} = a_{n_k}(h(x_{n_k}) + e_{n_k+1}) \xrightarrow[k\to\infty]{} 0$，但 $v(x_{n_k+1}) > \delta_1 \geqslant v(x_{n_k})$，由于 $v(\cdot)$ 连续，所以 $v(x_{n_k}) \xrightarrow[k\to\infty]{} \delta_1$.

由于 x_{n_k} 有界，可取收敛子列. 为简化符号，仍用 x_{n_k} 表示收敛子列，$x_{n_k} \xrightarrow[k\to\infty]{} \bar{x}$，即

$$v(\bar{x}) = \lim_{k\to\infty} v(x_{n_k}) = \delta_1. \tag{2.4.6}$$

所以 $d(\bar{x}, J) \triangleq \delta > 0$，取 T 充分小，由此及式(2.4.5)及 $x_{n_k} \to \bar{x}$ 知，对充分大的 k，

$$d(x_m, J) \geqslant \frac{\delta}{2}, \ \forall m: n_k \leqslant m \leqslant m(n_k, T). \tag{2.4.7}$$

用式(2.4.5)及微分中值公式，知存在 ξ: $\|\xi - x_{n_k}\| \leqslant c_1 T$，使

$$\begin{aligned} v(x_{m(n_k,T)+1}) - v(x_{n_k}) &= (x_{m(n_k,T)+1} - x_{n_k})^\tau v_x(\bar{x}) \\ &\quad + (x_{m(n_k,T)+1} - x_{n_k})^\tau (v_x(\xi) - v_x(\bar{x})). \end{aligned} \tag{2.4.8}$$

由于 $x_{n_k} \to \bar{x}$，所以当 k 充分大时，$\|\xi - \bar{x}\| \leqslant 2c_1 T$，所以从(2.4.5)知，当 $T \to 0$ 时，式(2.4.8)中的最后一式的量级为 $o(T)$，从引理假设，x_m 由式(2.1.2)及式(2.1.5)定义，$n_k \leqslant m \leqslant m(n_k, T)$，所以从式(2.4.8)知

$$\begin{aligned} v(x_{m(n_k,T)+1}) - v(x_{n_k}) &= \sum_{i=n_k}^{m(n_k,T)} a_i y_{i+1}^\tau v_x(\bar{x}) + o(T) \\ &= \sum_{i=n_k}^{m(n_k,T)} a_i h^\tau(x_i) v_x(x_i) + \sum_{i=n_k}^{m(n_k,T)} a_i h^\tau(x_i)(v_x(\bar{x}) \\ &\quad - v_x(x_i)) + \sum_{i=n_k}^{m(n_k,T)} a_i v_x^\tau(\bar{x}) e_{i+1} + o(T) \end{aligned} \tag{2.4.9}$$

从式(2.4.3)、式(2.4.5)及 $v_x(\cdot)$ 的连续性知，上式右端第二项为 $o(T)$，而用条件 A2.4.3 知第三项的范数当 $k \to \infty$ 时为 $o(T)$.

用式(2.4.5)及 $\|x_{n_k}\| \leqslant c$ 知

$\|x_i\| \leqslant (c + c_1 T)$，$\forall i$: $n_k \leqslant i \leqslant m(n_k, T)$，利用条件(2.4.2)及式(2.4.7)，从式(2.4.9)知存在 $\alpha > 0$ 及充分小的 $T > 0$，只要 k 充

分大就有

$$v(x_{m(n_k,T)+1})-v(x_{n_k})\leqslant-\alpha T, \tag{2.4.10}$$

所以

$$\limsup_{k\to\infty} v(x_{m(n_k,T)+1})\leqslant\delta_1-\alpha T. \tag{2.4.11}$$

另一方面，从式(2.4.5)知

$$\max_{n_k\leqslant m\leqslant m(n_k,T)+1}|v(x_m)-v(x_{n_k})|\xrightarrow[T\to 0]{}0. \tag{2.4.12}$$

由于 $v(x_{n_k})\leqslant\delta_1$，$v(x_{m_k})\geqslant\delta_2$，$\delta_1<v(x_i)<\delta_2$，$\forall i: n_k<i<m_k$，所以从式(2.4.12)知对充分小的 T 有

$$m(n_k,\ T)+1<m_k.$$

也就是说 $v(x_{m(n_k,T)+1})\in[\delta_1,\ \delta_2)$，

这和式(2.4.11)矛盾．这就证明了 i)．

ii) 当 $\{x_n\}$ 有界时，则有收敛子列 $x_{n_k}\to\bar{x}$，所以式(2.4.6)～式(2.4.11)照样成立．但式(2.4.11)左边的极限为 δ_1，所以得出矛盾的不等式．这就证明了引理． □

这个引理比下面定理 2.4.1 实际需要更广一些，我们把它叙述成现在这个形式，是为了使它对下一章也适用．

定理 2.4.1 设对某一 ω 条件 A2.4.0～A2.4.3 成立，并对取定的初值 x_0，由算法(2.1.5)定义的 x_n 一致有界，那么对这个 ω，$d(x_n,\ J)\xrightarrow[n\to\infty]{}0$．

证明 我们先证 $\lim\limits_{n\to\infty}d(v(x_n),\ v(J))=0$ a.s.． (2.4.13)

记

$$v_1\triangleq\liminf_{n\to\infty}v(x_n)\leqslant\limsup_{n\to\infty}v(x_n)\triangleq v_2. \tag{2.4.14}$$

1) 设 $v_1=v_2$，根据引理 2.4.1 之 ii)，必有 $d(v_1,\ v(J))=0$，所以 $\lim\limits_{n\to\infty}d(v(x_n),\ v(J))=0$．

2) 设 $v_1<v_2$，并设至少有一个 v_i，$i=1, 2$ 不属于 $\overline{v}(J)$，不失一般性，设 $v_1\overline{\in}\overline{v}(J)$，这里 $\overline{v}(J)$ 表示 $v(J)$ 的闭包．那么必存在 $\varepsilon>0$，使 $[v_1+\varepsilon,\ v_1+2\varepsilon]\cap\overline{v}(J)=\phi$，且 $v_1+2\varepsilon<v_2$．从 v_1，v_2 的定义(2.4.14)知，$v(x_n)$ 无穷次穿越区间 $[v_1+\varepsilon,\ v_1+2\varepsilon]$．根据引

理2.4.1之i)这不可能. 所以 v_1 及 v_2 均属 $\overline{v}(J)$, 也就是说式(2.4.13)成立.

现在我们利用条件 $A2.4.2$ 之 b),来证 $d(x_n, J)\to 0$.

1) 设 $v(x)=c_2$, $\forall x\in J$, 那么根据式(2.4.13)知 $v(x_n)\to c_2$. 反设存在子列 $x_{n_k}\to\bar{x}$, $d(\bar{x}, J)>0$. 记 $d(\bar{x}, J)=\delta>0$, 那么式(2.4.7)~式(2.4.10)照样成立. 由于当 $n\to\infty$ 时 $v(x_n)\to c_2$, 所以当 $k\to\infty$ 时, 式(2.4.10)左端趋于0, 从而得出矛盾的不等式, 因此 $d(x_n, J)\to 0$.

2) 设对任一 x, 只要 $d(x, J)>0$ 就有 $d(v(x), v(J))>0$, 又设 x_{n_k} 为 x_n 的任一收敛子列, $x_{n_k}\to\bar{x}$. 从式(2.4.13)知 $d(v(\bar{x}), v(J))=0$, 所以 $d(\bar{x}, J)=0$. 由于 x_{n_k} 的任意性, 便知 $d(x_n, J)\xrightarrow[n\to\infty]{}0$. □

从这一章的引理及定理看, 在证明算法收敛性时, 我们充分利用了 Lyapunov 函数的性质, 而没有引进插值函数及它的尾函数所满足的微分方程.

下面的定理将适当加强对 $h(\cdot)$ 的要求, 但去掉 Lyapunov 函数的存在性条件 A 2.4.2.

定理 2.4.2 设 $f(\cdot)$ 是 $R^l\to R$ 的二次连续可微函数, $\sup\limits_{x\in R^l}f(x)\leqslant f(x)$, $x\in J$, J 为有界集($f(x)=\text{const}$, $\forall x\in J$), 或记 $h(x)=f_x(x)$, $h(x)=0$, $\forall x\in J$. 还设成立条件 A2.4.1 及 A 2.4.3, 且对取定的初值 x_0, 由算法(2.1.5)定义的 x_n 有界, 那么 $d(x_n, J)\xrightarrow[n\to\infty]{}0$.

证明 只要证明存在满足 A 2.4.2 的 $v(x)$, 那么利用定理 2.4.1 就得到所要证的结论. 我们取 $v(\cdot)=-f(\cdot)$, 那么 $h^\tau(x)v_x(x)=-\|h(x)\|^2$, 由于 $v(x)$ 在 J 上取常值, 所以 A 2.4.2 成立. □

定理 2.4.3 在定理 2.4.1 及 2.4.2 中把 A 2.4.1 加强为

A2.4.1′ a_n 非升, $a_n>0$, $a_n\to 0$, $\sum\limits_{n=1}^{\infty}a_n=\infty$, 并且存在常数 μ, 使

$$a_{n+1}^{-1}-a_n^{-1}\leqslant\mu,\ \forall n\geqslant 1. \tag{2.4.15}$$

把 A2.4.3 改为

A2.4.3′: $\varepsilon_n=e_n+\nu_n a_n\sum_{i=1}^{n}e_{i+1}\to 0,\ \nu_n\to 0$, 而保持其它条件不变, 那么这两个定理的结论照样成立.

证明　由于$\nu_n\xrightarrow[n\to\infty]{}0$, 所以如果我们能够证明

$$\sum_{i=n_k}^{m}a_ie_{i+1}\xrightarrow[k\to\infty]{}0,\ \forall m:\ n_k\leqslant m\leqslant m(n_k,\ T), \tag{2.4.16}$$

那么 A2.4.3 成立, 也就保证了定理成立.

当 A2.4.3′ 成立时, 用式(2.4.15)知, 对 m　$n_k\leqslant m\leqslant m(n_k, T)$ 有

$$\left\|\sum_{i=n_k}^{m}a_ie_{i+1}\right\|=\left\|a_m\sum_{i=1}^{m}e_{i+1}-a_{n_k}\sum_{i=1}^{n_k-1}e_{i+1}+\sum_{i=n_k}^{m-1}(a_i-a_{i+1})\sum_{j=1}^{i}e_{j+1}\right\|$$
$$\leqslant a_m\left\|\sum_{i=1}^{m}e_{i+1}\right\|+a_{n_k-1}\left\|\sum_{i=1}^{n_k-1}e_{i+1}\right\|+\mu\sum_{i=n_k}^{m-1}a_{i+1}\left\|a_i\sum_{j=1}^{i}e_{j+1}\right\|\xrightarrow[k\to\infty]{}0. \tag{2.4.17}$$

□

第 3 章

随机变界截尾算法

在第 2 章中，我们已经讨论 Robbins-Monro 算法，给出各种分析收敛性的方法. 用鞅方法时，对回归函数和量测噪声要求都比较严. 用 ODE 方法时，对量测噪声有较大放松，但对回归函数的增长速度仍有要求，并且还事先假定估计值对时间有界. 用 Lyapunov 函数方法时，与 ODE 方法相比，对回归函数的要求有进一步放松. 在这一章中，我们对 RM 算法的有限步作一些修正后，不再要求算法的有界性，对回归函数和量测噪声的要求都较弱，用 Lyapunov 函数法来证算法的收敛性，并给出收敛速度和稳健性分析.

§3.1 随机变界截尾算法的收敛性分析

在第 2 章的 RM 算法中，我们看到 ODE 方法可以减弱对量测误差的限制，用 Lyapunov 函数方法进一步可减弱对回归函数的要求，但要预先假设 $\{x_n\}$ 一致有界. 在 § 2.2 中，用鞅方法可以证明 $\{x_n\}$ 一致有界，但对量测误差的要求很强，而且回归函数必须满足如式(2.1.12)之类的线性增长条件. 现在去掉对 $\{x_n\}$ 一致有界的假设以及对回归函数线性增长的要求.

容易发现，在原有的算法(2.1.7)的框架下，这些条件很难去掉. 先看一个简单的例子.

例 3.1.1 设 $h(x)=-x^3$, $\varepsilon_n\equiv 0$, $a_n=\dfrac{1}{n+1}$, $\nu(x)=x^2$. 容易验证，此例满足 A2.3, $x^0=0$, 若取 $x_0=10$, 由算法 $x_{n+1}=x_n+$

$\frac{1}{n+1}(-x_n^3)$，可以得到

$$x_1=-990,\ x_2=485148610,\ x_3\doteq-3.9\times10^{24}.$$

仅由以上三步递推，已经可以看出 x_n 将迅速发散．说明即使对无量测误差的特殊情况，只要回归函数不具有线性增长的性质，算法(2.1.7)都不能保证对初值的大范围收敛性．

但是，在上例中，如果取 $|x_0|<1$，容易验证，x_n 将收敛到 $x^0=0$．这说明，当回归函数的非线性较强时，初值如果偏差 x^0 较远，回归函数的值将很大，而步长因子列$\{a_n\}$的衰减速度相对较慢，不足以压制住这种影响，才使 x_n 迅速向无穷发散．如果能设法使回归函数的增长速度降下来，有可能使算法大范围收敛．

另一方面，在实际问题中，要处理的回归函数往往都具有较强的非线性，而标准的 RM 算法对这类情况又不具有大范围收敛性．于是，人们将标准算法改造成投影算法或截尾算法[49,102]，如

$$x_{n+1}=\pi_S(x_n+a_ny_{n+1}),\tag{3.1.1}$$

其中 S 是零点所在的凸集，π_S 是对 S 的投影算子．或者

$$x_{n+1}=\begin{cases}x_n+a_ny_{n+1}, & \text{如果 } x_n+a_ny_{n+1}\in S,\\ x^*, & \text{如果 } x_n+a_ny_{n+1}\overline{\in} S,\end{cases}\tag{3.1.2}$$

其中 x^* 为 S 中的任一点．

这种投影或截尾算法实质上是用构造的办法来实现算法的有界性．由于需要 x^0 的所在范围 S 的先验信息，当然不是具有全局收敛性的算法．

我们设想，如果式(3.1.2)中的 S 不是一成不变的，而是每发生一次截尾就让它变大，这样若把每一次递推时所用的 S 写成序列$\{S_n\}$，$S_i\subset S_j$，$\forall j>i$，只要截尾发生无穷多次，S_n 就趋于无穷．那么，任取 S_0，即使 x^0 不包含在 S_0 内，从某一次 N 开始，x^0 总含在 S_N 内，因而也在 S_n 内，$n\geqslant N$．但是，即使 $x^0\in S_N$，并不能保证 $x_n+a_ny_{n+1}$ 不再跑到 S_n 外，$n\geqslant N$，如果这种情况发生无穷多次，算法显然不收敛．因此，这种"变截尾"的新构想能否成功，就在于能否证明在 S_n 套住 x^0 以后，截尾至多只进行有限多次．由于截尾使

x_n 将多次回到某一固定的有界集内，而步长因子 a_n 却在不断变小，这相当于使 $h(x_n)$ 的增长速度受到压制，在等待 a_n 变得足够小，直至可以压制住 x_n 偏离零点较远时 $h(x_n)$ 在算法中的作用. 在例3.1.1中，虽然 $x_0=10$，但经过多次截尾后，由于 $\frac{1}{n}$ 已变得很小，而 x_n 不断回到有界集，使 $\frac{1}{n}x_n^3$ 也很小，相当于取 $|x_0|<1$ 的情况.

现在我们用数学语言来精确描述随机变界截尾后的RM算法.

设 $\{M_n\}$ 是一个单调趋于无穷的正实数序列，x^* 为 R^l 中某一点. 递推地定义一个正整数值随机序列 $\{\sigma_n\}$ 和算法 $\{x_n\}$ 如下：

$$\sigma_n=\sum_{i=0}^{n-1} I_{[|x_i+a_i y_{i+1}|>M_{\sigma_i}]},\quad \sigma_0=0, \tag{3.1.3}$$

$$x_{n+1}=(x_n+a_n y_{n+1}) I_{[|x_n+a_n y_{n+1}|\leq M_{\sigma_n}]}+x^* I_{[|x_n+a_n y_{n+1}|>M_{\sigma_n}]}, \tag{3.1.4}$$

和算法(3.1.2)相比，这里的 S 简单地取作以原点为中心的球. 在以后算法收敛性的证明中可以看到，若回到式(3.1.2)那种更一般的框架并不影响证明，只是现在的简化可以使以后的叙述更简明. 此外，也可以类似设计一种"变投影"的算法，只是每次投影后 x_n 不是回到一固定点而是回到一逐渐变大的 S 的边界上，因而为控制住 $h(x_n)$ 的增长速度，应让 S 变大的速度大大低于 $\{a_n\}$ 趋于0的速度才能保证算法的收敛性.

下面，给出使本节一开始所提到的希望得以实现的算法全局收敛性定理.

定理3.1.1[30] 设对某一 ω 成立条件A2.4.0～A2.4.3，还设存在常数 c_0 使 $v(x^*)<\inf_{|x|=c_0} v(x)$，同时 $\|x^*\|<c_0$，并且

$$(\overline{v}(J))^c\cap\left(v(x^*),\ \inf_{|x|=c_0} v(x)\right)\neq\varnothing, \tag{3.1.5}$$

则对任意初值 x_0，算法(3.1.3)、(3.1.4)所定义的 x_n 在这个 ω 上有 $\lim_{n\to\infty} d(x_n, J)=0$.

注 3.1.1 和定理 2.4.1 相比，除了算法不同外，这里少了一个$\{x_n\}$一致有界的条件.

注 3.1.2 当 A2.4.2 之 b) 中 $v(J)=\text{const}$ 时，式(3.1.5)显然成立. 当 $v(J)$不稠于任一区间时，式(3.1.5)也一定成立.

我们先证引理，都是针对固定样本 ω 的.

引理 3.1.1 设 A 2.4.1 和 A 2.4.3 成立，$\{x_{n_k}\}$是一个由算法(3.1.3)、(3.1.4)定义的$\{x_n\}$的收敛子序列. 则存在常数 $M>0$, $\Delta>0$ 和 $k_T>0$, 使得对任意 $k>k_T$, $T\in[0, \Delta)$和 m: $n_k\leqslant m\leqslant m(n_k, T)$有

$$\left\|\sum_{i=n_k}^{m+1} a_i y_{i+1}\right\|\leqslant M, \qquad (3.1.6)$$

其中 $m(n, T)$由式(2.4.3)定义.

证明 若算法(3.1.3)、(3.1.4)中截尾至多发生有限次，即存在 N, 使 $\sigma_n=\sigma_N$, $\forall n\geqslant N$. 则算法一致有界，显然只要 $n_k\geqslant N$ 就有

$$\left\|\sum_{i=n_k}^{m+1} a_i y_{i+1}\right\|=\|x_{m+2}-x_{n_k}\|\leqslant 2M_{\sigma_N}, \qquad (3.1.7)$$

这时可取 $M=2M_{\sigma_N}$.

现设 $\sigma_n\xrightarrow[n\to\infty]{}\infty$, $\bar{x}=\lim\limits_{k\to\infty} x_{n_k}$. 反设引理不成立. 取实数 $c>\|\bar{x}\|$. 存在 k_c, 只要 $k\geqslant k_c$ 就有

$$\|x_{n_k}\|\leqslant(c+\|\bar{x}\|)/2 \qquad (3.1.8)$$

取 $T_s>0$, $T_s\xrightarrow[s\to\infty]{}0$. 据反假设，对任何 s, 存在 k_s 和 m_s. $n_{k_s}\leqslant m_s\leqslant m(n_{k_s}, T_s)$使得

$$\left\|\sum_{i=n_{k_s}}^{m_s+1} a_i y_{i+1}\right\|>(c-\|\bar{x}\|)/2 \qquad (3.1.9)$$

不失一般性，可认为 $k_s>k_{s-1}>k_c$, $\forall s\geqslant 1$, 并设

$$m_s=\inf\left\{m:\ \left\|\sum_{i=n_{k_s}}^{m+1} a_i y_{i+1}\right\|>(c-\|\bar{x}\|)/2\right\}, \qquad (3.1.10)$$

那么对 m: $n_{k_s}\leqslant m\leqslant m_s$, 由式(3.1.8)和(3.1.10)知

$$\left\|x_{n_{k_s}}+\sum_{i=n_{k_s}}^{m} a_i y_{i+1}\right\|\leqslant c. \qquad (3.1.11)$$

从 $\sigma_n \xrightarrow[n\to\infty]{} \infty$ 知，存在 s_0，当 $s\geqslant s_0$ 时，$M_{\sigma_{n_{k_s}}}>c$. 由式(3.1.11)知

$$\left\|x_{n_{k_s}}+\sum_{i=n_{k_s}}^{m} a_i y_{i+1}\right\|\leqslant M_{\sigma_{n_{k_s}}},\ \forall m;\ n_{k_s}\leqslant m\leqslant m_s \quad (3.1.12)$$

因而由式(3.1.4)可得:

$$x_{m+1}=x_m+a_m y_{m+1},\ \forall m;\ n_{k_s}\leqslant m\leqslant m_s, \quad (3.1.13)$$

再由式(3.1.8)、(3.1.11)和(3.1.13)知

$$\|x_m\|\leqslant c,\ \|h(x_m)\|\leqslant c',\ \forall m;\ n_{k_s}\leqslant m\leqslant m_s+1, \quad (3.1.14)$$

对任意 $T>0$, 只要 s 充分大，必有 $T_s<T$ 及 $m_s<m(n_{k_s}, T)$. 从式(2.4.4)知对任意 $T>0$,

$$\limsup_{s\to\infty}\left\|\sum_{i=n_{k_s}}^{m_s} a_i\varepsilon_{i+1}\right\|=o(T).$$

由于 T 可任意小, 所以

$$\limsup_{s\to\infty}\left\|\sum_{i=n_{k_s}}^{m_s} a_i\varepsilon_{i+1}\right\|=0,\ \lim_{s\to\infty}\left\|\sum_{i=n_{k_s}}^{m_s+1} a_i\varepsilon_{i+1}\right\|=0,\ \lim_{s\to\infty} a_{m_s+1}\varepsilon_{m_s+2}=0. \quad (3.1.15)$$

用式(3.1.14)及式(3.1.15)便知

$$\left\|x_{m_s+1}-x_{n_{k_s}}\right\|\leqslant \sum_{i=n_{k_s}}^{m_s} a_i\|h(x_i)\|+\left\|\sum_{i=n_{k_s}}^{m_s} a_i\varepsilon_{i+1}\right\|\xrightarrow[s\to\infty]{} 0, \quad (3.1.16)$$

和

$$\|x_{m_s+1}-x_{n_{k_s}}+a_{m_s+1}y_{m_s+2}\|\leqslant\|x_{m_s+1}-x_{n_{k_s}}\|+\|a_{m_s+1}y_{m_s+2}\|\xrightarrow[s\to\infty]{}0, \quad (3.1.17)$$

上式右端趋于 0, 是因为最后一项

$$\|a_{m_s+1}y_{m_s+2}\|\leqslant a_{m_s+1}\|h(x_{m_s+1})\|+\|a_{m_s+1}\varepsilon_{m_s+2}\|\xrightarrow[s\to\infty]{} 0.$$

另一方面, 从式(3.1.10)知

$$\|x_{m_s+1}-x_{n_{k_s}}+a_{m_s+1}y_{m_s+2}\|=\left\|\sum_{i=n_{k_s}}^{m_s+1} a_i y_{i+1}\right\|>(c-\|\bar{x}\|)/2,$$

这与式(3.1.17)矛盾, 因而引理成立. □

引理 3.1.2 在引理 3.1.1 的条件下, 成立估计

$$\|x_{m+1}-x_{n_k}\|\leqslant c_1T,\ \forall m:n_k\leqslant m\leqslant m(n_k,\ T),\ \forall k\geqslant k_T. \tag{3.1.18}$$

证明 首先考虑 $\lim\limits_{n\to\infty}\sigma_n=\infty$ 的情况. 显然,这时存在常数 k_0 和 $c_2>0$,使得对 $\forall k\geqslant k_0,\ n\geqslant n_k$,

$$M_{\sigma_{n_k}}>M+1+\|\bar{x}\|,\ \|x_{n_k}\|\leqslant 1+\|\bar{x}\|, \tag{3.1.19}$$

这里的 M 和 $\bar{x}$ 如引理 3.1.1 中定义.

于是由引理 3.1.1 和式(3.1.19),存在 $k_T\geqslant k_0$, $\Delta>0$, 使得对任 $T\in(0,\ \Delta)$和 $m:n_k\leqslant m\leqslant m(n_k,\ T)$,只要 $k\geqslant k_T$,有

$$\left\|x_{n_k}+\sum_{i=n_k}^{m}a_iy_{i+1}\right\|\leqslant M+1+\|\bar{x}\|\leqslant M_{\sigma_{n_k}}, \tag{3.1.20}$$

因而

$$x_{m+1}=x_m-a_my_{m+1}, \tag{3.1.21}$$

$$\|x_{m+1}\|\leqslant M+1+\|\bar{x}\|,\ \|h(x_m)\|\leqslant c_3,\ \forall m:n_k\leqslant m\leqslant m(n_k,\ T). \tag{3.1.22}$$

再由式(3.1.22)可得:

$$\left\|\sum_{i=n_k}^{m}a_ih(x_i)\right\|\leqslant c_3\sum_{i=n_k}^{m}a_i\leqslant c_3T. \tag{3.1.23}$$

而由 A 2.3.3 易知,不失一般性,当 $k\geqslant k_T$,

$$\left\|\sum_{i=n_k}^{m}a_ie_{i+1}\right\|\leqslant c_4T, \tag{3.1.24}$$

只要令 $c_1=c_3+c_4$,则由式(3.1.23)和(3.1.24)推出式(3.1.18).

当 $\lim\limits_{n\to\infty}\sigma_n=M_2<\infty$, 则以上推理不需引理 3.1.1 都可进行. 引理得证. □

定理 3.1.1 的证明.

根据定理假设(3.1.5), 知存在 $\delta_1\in\left(v(x^*),\ \inf\limits_{\|x\|=c_0}v(x)\right)$, 使 $d(\delta_1,\ v(J))>0$,所以存在 $\delta_2>\delta_1$,使 $d([\delta_1,\ \delta_2],\ v(J))>0$.

下面来证 $\lim\limits_{n\to\infty}\sigma_n=\sigma<\infty$.

假设对某个 ω, $\sigma=\infty$,那么 x_n 从 x^* 出发将无穷次穿越球面 $\{x:\|x\|=c_0\}$,

注意到$\|x^*\|<c_0$，所以存在子列$\{x_{n_k}\}$及$\{x_{m_k}\}$，$n_k<m_k$，$\delta_1<v(x_i)<\delta_2$，$\|x_i\|<c_0$，$\forall i$：$n_k<i<m_k$，$v(x_{n_k})\leqslant\delta_1$，$v(x_{m_k})\geqslant\delta_2$，也就是说$v(x_n)$无穷次穿越区间$[\delta_1,\delta_2]$，并且$\|x_{n_k}\|$有界. 从引理3.1.2及式(3.1.12)知引理2.4.1的条件成立. 根据该引理知，这不可能发生，所以$\sigma<\infty$，即存在依赖于ω的n_0，当$n\geqslant n_0$时，$\sigma_n\equiv\sigma$. 因此$\{x_n\}$有界，并且当$n\geqslant n_0$时，算法(3.1.4)和RM算法完全重合. 下面的证明完全和定理2.4.1的证明相同，便知定理结论正确. □

定理3.1.2 设$f(\cdot)$是$R^l\to R$的二次连续可微函数，J为$f(\cdot)$的极值点集，$f(x)=\text{const}$，$x\in J$. 记$h(x)=f_x(x)$. 还设A2.4.1及A2.4.3成立，并且存在常数c_0，使$f(x^*)>\sup\limits_{\|x\|=c_0}f(x)$，那么对任意初值$x_0$，算法(3.1.3)、(3.1.4)所定义的$x_n$ a.s.收敛到J，即

$$d(x_n,J)\xrightarrow[n\to\infty]{}0\quad\text{a.s.}$$

证明 取$v(\cdot)=-f(\cdot)$，可知条件A2.4.2成立，并且条件(3.1.5)显然成立. 所以定理由定理3.1.1即得 □

定理3.1.3 如果把定理3.1.1及3.1.2中的条件A2.4.1及A2.4.3换成A2.4.1′及A2.4.3′，而把其余条件保持不变，那么定理结论仍成立.

证明 和定理2.4.3的证明相同. □

定理3.1.2是求回归函数的极值，但要求量测到回归函数的梯度，尽管可能带有误差. 在有些实际问题中，不能量测回归函数的梯度，而只能带误差地量测回归函数本身. 下面就是讨论这种情形.

设回归函数$h(\cdot)$是$R^l\to R$的连续可微实值函数. $h_x(x)=0$，$\forall x\in J$.

记
$$x_n=[x_n^1,\ \cdots,\ x_n^l]^\tau,$$
$$x_n^{i+}=[x_n^1,\ \cdots,\ x_n^{i-1},\ x_n^i+b_n,\ x_n^{i+1},\ \cdots,\ x_n^l]^\tau,$$
$$x_n^{i-}=[x_n^1,\ \cdots,\ x_n^{i-1},\ x_n^i-b_n,\ x_n^{i+1},\ \cdots,\ x_n^l]^\tau,\ 1\leqslant i\leqslant l,\quad(3.1.25)$$

$$\nabla h(x_n)=\frac{1}{2bn}[h(x_n^{1+})-h(x_n^{1-}),\ \cdots,\ h(x_n^{l+})-h(x_n^{l-})]^{\tau},$$

(3.1.26)

其中$\{b_n\}$为标量序列，满足

$$b_n>0,\ b_n\xrightarrow[n\to\infty]{}0. \qquad (3.1.27)$$

现在，我们观测

$$y_{n+1}=\nabla h(x_n)+\varepsilon_{n+1},\ n>0,\ y_0=0. \qquad (3.1.28)$$

显然$y_{n+1}\in R^l$，是由对两个l维矢量

$$[h(x_n^{1+}),\ \cdots,\ h(x_n^{l+})]^{\tau},\ [h(x_n^{1-}),\ \cdots,\ h(x_n^{l-})]^{\tau} \qquad (3.1.29)$$

的观测值相减后乘以$\frac{1}{2b_n}$得到的，这相当于在量测回归函数的差商．因此，这里不仅有随机量测误差，还有差商与导数之间的误差．

我们现在采用的算法是式(3.1.28)、(3.1.3)和(3.1.4)，称为变界截尾 Kiefer-Wolfowitz(KW)算法．与变界截尾 RM 算法相比，只是在式(3.1.3)、(3.1.4)中的y_{n+1}定义如式(3.1.28)．

引进如下的假设组 A 3.1：

A 3.1.0　在任何有界集上，$h_x(\cdot)$满足 Lipschitz 条件．

A 3.1.1　和 A 2.3.1 相同．

A 3.1.2　和 A 2.4.2 相同，只是把那里的$h(\cdot)$换成$h_x(\cdot)$．

A 3.1.3　和 A 2.3.3 相同．

于是有变界截尾 KW 算法的大范围收敛性定理如下：

定理 3.1.4[30]　在假设 A 3.1 下，对任意初值x_0，算法(3.1.28)、(3.1.3)和(3.1.4)给出的x_n趋于J，即

$$d(x_n,\ J)\xrightarrow[n\to\infty]{}0\quad \text{a.s.}$$

证明　重写式(3.1.28)为

$$y_{n+1}=h_x(x_n)+[\nabla h(x_n)-h_x(x_n)]+e_{n+1}+\nu_{n+1}, \qquad (3.1.30)$$

由条件 A3.1.0 以及中值定理，$b_n\xrightarrow[n\to\infty]{}0$，易见

$$|\nabla h(x_n)-h_x(x_n)|\xrightarrow[n\to\infty]{}0 \qquad (3.1.31)$$

因而若把 $\nabla h(x_n)-h_x(x_n)$ 也并入 ν_{n+1}，式(3.1.30)中的 y_{n+1} 也可看作是对 $h_x(x_n)$ 的量测，其量测误差仍满足 A 3.1.3，且算法(3.1.28)、(3.1.3)和 (3.1.4) 也可看成是量测 $h_x(\cdot)$ 的变界截尾 RM 算法. 定理 3.1.1 的条件均得到满足，故定理成立. □

从以上证明过程中可以看出，只要假设 A 3.1.0 成立，并采用量测(3.1.25)～(3.1.28)，就相当于对回归函数的梯度 $h_x(\cdot)$的观测，问题化为 RM 算法. 因此，在以下各节中，我们不再讨论 KW 算法，只要研究 RM 算法就够了.

§3.2 收敛速度

这一节中我们给出算法(3.1.3)、(3.1.4)给出的 x_n 的收敛速度.

以下设 $h(\cdot)$有唯一零点 x^0，即 $J=x^0$. 先讨论强收敛速度，为此引入条件 A 3.2：

A 3.2.0 $R^l\to R^l$ 的回归函数 $h(\cdot)$在任一有界集上有界，并当 $x\to x^0$ 时，

$$h(x)=H(x-x^0)+\delta(x),\ \delta(x^0)=0,\ \delta(x)=o(\|x-x^0\|), \tag{3.2.1}$$

H 为稳定阵，并且

$$H+\alpha\delta I \text{ 也稳定 （本征值实部为负）} \tag{3.2.2}$$

α 和 δ 的定义见下面式(3.2.4)及式(3.2.5).

A 3.2.1 确定性步长因子$\{a_n\}$满足以下条件

$$a_n>0,\ a_n\xrightarrow[n\to\infty]{}0,\ \sum_{n=1}^{\infty}a_n=\infty, \tag{3.2.3}$$

$$\frac{a_n-a_{n+1}}{a_na_{n+1}}\xrightarrow[n\to\infty]{}\alpha\geqslant 0. \tag{3.2.4}$$

A 3.2.2 量测误差列$\{\varepsilon_n\}$表达为 $\varepsilon_n=e_n+\nu_n$，对某 δ：$0<\delta\leqslant 1$，$\nu_{n+1}=O(a_n^\delta)$ a.s.，并且

$$\sum_{n=1}^{\infty}a_n^{1-\delta}e_{n+1}\ \text{a.s. 收敛.} \tag{3.2.5}$$

A 3.2.3 存在二次连续可微函数 $v(\cdot)$: $R^l \to R$, 使对任意 $\Delta_2 > \Delta_1 > 0$,

$$\sup_{\Delta_1 \leq \|x - x^0\| \leq \Delta_2} h^\tau(x) v_x(x) < 0, \tag{3.2.6}$$

还设存在常数 c_0 使 $v(x^*) < \inf\limits_{\|x\| = c_0} v(x)$, $\|x^*\| < c_0$, x^* 定义在式(3.1.4)中.

注 3.2.1 i) 若 $a_n = \frac{a}{n}$, $a>0$, 则易见它满足 A 3.2.1, 这时 $\alpha = \frac{1}{a}$. 同时式(3.2.2)应为 $H + \frac{\delta}{a} I$ 稳定. ii) 若 $a_n = \frac{a}{n^{\frac{1}{2}+\beta}}$, $a>0$, $\beta>0$, 则可验证它满足式(3.2.1), 这时 $\alpha = 0$, 式(3.2.2)等价于 H 稳定. 设$\{e_n\}$为条件二阶矩有界的鞅差列, 即

$$E(e_n | \mathscr{F}_{n-1}) = 0, \ E(\|e_n\|^2 | \mathscr{F}_{n-1}) < M, \tag{3.2.7}$$

当 $\delta>0$ 充分小时, $\beta - \delta\left(\frac{1}{2} + \beta\right) > 0$, 那么 $\sum\limits_{n=1}^{\infty} a_n^{2(1-\delta)} < \infty$, 所以利用鞅收敛定理(定理 1.2.2)立即可知式(3.2.5)成立. 同样, 当$\{e_n\}$为有限项鞅差列所构成的滑动平均(MA)列, 式(3.2.5)也成立.

为了给出算法的强收敛速度, 需要引入下面的引理[103].

引理 3.2.1 设 $l \times l$ 维方阵 $H_n \xrightarrow[n\to\infty]{} H$, H 为稳定阵, 数列$\{a_n\}$满足式(3.2.3), l 维矢量列$\{e_n\}$, $\{\nu_n\}$满足

$$\sum_{n=1}^{\infty} a_n e_{n+1} \text{收敛}, \nu_n \xrightarrow[n\to\infty]{} 0, \tag{3.2.8}$$

则如下定义的矢量列$\{x_n\}$

$$x_{n+1} = x_n + a_n H_n x_n + a_n (e_{n+1} + \nu_{n+1}) \tag{3.2.9}$$

对任意初值趋于 0.

证明 记

$$\Phi_{n,j} \triangleq (I + a_n H_n) \cdots (I + a_j H_j), \ \Phi_{j,j+1} \triangleq I. \tag{3.2.10}$$

我们来证存在常数 $c_0>0$, $c>0$, 使

$$\|\Phi_{n,j}\| \leq c_0 \exp\left[-c \sum_{k=j}^{n} a_k\right], \ \forall n \geq j, \ \forall j \geq 0. \tag{3.2.11}$$

因为 H 稳定，存在正定阵 P，使

$$PH+H^{\tau}P=-2I.$$

由于 $H_n \xrightarrow[n\to\infty]{} H$，所以存在 n_0，当 $j\geqslant n_0$ 时

$$PH_n+H_n^{\tau}P\leqslant -I. \tag{3.2.12}$$

那么，

$$\begin{aligned}\Phi_{n,j}^{\tau}P\Phi_{n,j}&\leqslant\Phi_{n-1,j}^{\tau}(P-a_nI+a_n^2H^{\tau}PH)\Phi_{n-1,j}\\&=\Phi_{n-1,j}^{\tau}P^{\frac{1}{2}}(I-a_nP^{-1}+a_n^2P^{-\frac{1}{2}}H^{\tau}PHP^{-\frac{1}{2}})P^{\frac{1}{2}}\Phi_{n-1,j}.\end{aligned} \tag{3.2.13}$$

由于 $a_n\to 0$，所以存在 $c>0$，对一切充分大的 n（不妨认为 n_0 已充分大，$n\geqslant n_0$），

$$\|I-a_nP^{-1}+a_n^2P^{-\frac{1}{2}}H^{\tau}PHP^{-\frac{1}{2}}\|\leqslant 1-2ca_n<e^{-2ca_n}, \tag{3.2.14}$$

那么从式(3.2.13)、(3.2.14)知对 $\forall n\geqslant j$，$\forall j\geqslant n_0$，

$$\Phi_{n,j}^{\tau}P\Phi_{n,j}\leqslant\exp\left[-2c\sum_{k=j}^{n}a_k\right]I,\quad \|\Phi_{n,j}\|\leqslant\lambda_{\min}^{-\frac{1}{2}}(P)\exp\left[-c\sum_{k=j}^{n}a_k\right], \tag{3.2.15}$$

这里 $\lambda_{\min}(P)$ 表示 P 的最小本征值.

注意到对 $\forall j<n_0$，

$$\|\Phi_{n_0-1,j}\|\leqslant\prod_{k=j}^{n_0-1}(1+a_k\|H_k\|)\leqslant\prod_{k=1}^{n_0-1}(1+a_k\|H_k\|),$$

所以由式(3.2.15)知

$$\begin{aligned}\|\Phi_{n,j}\|&\leqslant\|\Phi_{n,n_0}\|\cdot\|\Phi_{n_0-1,j}\|\\&\leqslant\lambda_{\min}^{-\frac{1}{2}}(P)\exp\left[-c\sum_{k=n_0}^{n}a_k\right]\prod_{k=0}^{n_0-1}(1+a_k\|H_k\|)\\&\leqslant\lambda_{\min}^{-\frac{1}{2}}(P)\exp\left[c\sum_{k=j}^{n_0-1}a_k\right]\prod_{k=0}^{n_0-1}(1+a_k\|H_k\|)\exp\left[-c\sum_{k=j}^{n}a_k\right],\end{aligned}$$

这个不等式显然包含了式(3.2.15)，所以式(3.2.11)成立.

从式(3.2.9)知

$$x_{n+1}=\Phi_{n,0}x_0+\sum_{j=0}^{n}\Phi_{n,j+1}a_j(e_{j+1}+\nu_{j+1}), \tag{3.2.16}$$

我们来证它右端趋于0．首先由式(3.2.3)及式(3.2.11)知上式右端第一项趋于0，现估计第二项．

设对 $n\geqslant n_1$，$\|\nu_n\|<\varepsilon$，

$$\left\|\sum_{j=0}^{n}\Phi_{n,j+1}a_j\nu_{j+1}\right\|\leqslant c_0\sum_{j=0}^{n_1-1}\left[\exp\left(-c\sum_{k=j+1}^{n}a_k\right)\right]a_j\|\nu_{j+1}\|$$
$$+\varepsilon c_0\sum_{j=n_1}^{n}\exp a_j\left[-c\sum_{k=j+1}^{n}a_k\right],\tag{3.2.17}$$

上式右端第一项，当 $n\to\infty$ 时显然趋于0，而当 n_1 充分大时，对第二项可估计如下：

$$\varepsilon c_0\sum_{j=n_1}^{n}\exp a_j\left[-c\sum_{k=j+1}^{n}a_k\right]\leqslant 2\varepsilon c_0\sum_{j=n_1}^{n}\exp\left(a_j-\frac{ca_j^2}{2}\right)\left[-c\sum_{k=j+1}^{n}a_k\right]$$
$$\leqslant 2\varepsilon c_0\sum_{j=n_1}^{n}\exp(1-e^{-ca_j})\left[-c\sum_{k=j+1}^{n}a_k\right]\leqslant 2c_0\varepsilon,\tag{3.2.18}$$

所以式(3.2.17)右端第二项当 $n\to\infty$，$\varepsilon\to0$ 时趋于0．现在来估计 $\sum_{j=0}^{n}\Phi_{n,j+1}a_je_{j+1}$

记　$s_n=\sum_{j=0}^{n}a_je_{j+1}$，$s_{-1}=0$．根据式(3.2.18)，$s_n\to s<\infty$ a.s.．所以当 $j>n_2>n_1$ 时，$|s_j-s|\leqslant\varepsilon$，有

$$\begin{aligned}\sum_{j=0}^{n}\Phi_{n,j+1}a_je_{j+1}&=\sum_{j=0}^{n}\Phi_{n,j+1}(s_j-s_{j-1})\\&=s_n-\sum_{j=0}^{n}(\Phi_{n,j+1}-\Phi_{n,j})s_{j-1}\\&=s_n-\sum_{j=0}^{n}(\Phi_{n,j+1}-\Phi_{n,j})s\\&\quad-\sum_{j=0}^{n}(\Phi_{n,j+1}-\Phi_{n,j})(s_{j-1}-s)\\&=s_n-s+\Phi_{n,0}s-\sum_{j=0}^{n_2}(\Phi_{n,j+1}-\Phi_{n,j})(s_{j-1}-s)\\&\quad+\varepsilon\sum_{j=n_2+1}^{n}c_0a_j\|H\|\exp\left[-c\sum_{k=j+1}^{n}a_k\right],\end{aligned}$$

注意到式(3.2.18)，便知上式最后一项 $\leqslant 2c_0\|H\|\varepsilon$，所以

$$\sum_{j=0}^{n}\Phi_{n,j+1}a_je_{j+1}\xrightarrow[n\to\infty]{}0\quad\text{a.s.}.$$

这样就证明了 $x_n \xrightarrow[n\to\infty]{} 0$ a.s.. □

定理3.2.1 设条件 A 3.2.0～A 3.2.3 成立，则变界截尾算法(3.1.3)、(3.1.4)给出的$\{x_n\}$有如下收敛速度

$$\|x_n - x^0\| = o(a_n^\delta) \quad \text{a.s.}, \tag{3.2.19}$$

证明 不失一般性，设$x^0=0$，由于定理3.1.1的条件成立，所以$\lim\limits_{n\to\infty} x_n = 0$ a.s.，并且有限步后，算法就不再截尾，成了通常的RM算法.

由式(3.2.4)知，$\dfrac{a_n - a_{n+1}}{a_{n+1}} \xrightarrow[n\to\infty]{} 0$，利用 Taylor 展开式有

$$\left(\frac{a_n}{a_{n+1}}\right)^\delta = \left(1+\frac{a_n - a_{n+1}}{a_{n+1}}\right)^\delta = 1+\delta\,\frac{a_n - a_{n+1}}{a_{n+1}} + O\left(\left(\frac{a_n - a_{n+1}}{a_{n+1}}\right)^2\right), \tag{3.2.20}$$

另外，$\delta(x)$可写为

$$\delta(x) = \left(\delta(x)\frac{x^\tau}{\|x\|^2}\right)x, \tag{3.2.21}$$

或者

$$\delta(x_n) = D_n x_n,\ D_n = \left(\delta(x_n)\cdot\frac{x_n^\tau}{\|x_n\|^2}\right), \tag{3.2.22}$$

由式(3.2.1)知，$D_n \xrightarrow[n\to\infty]{} 0$.

于是由式(3.2.1)、(3.2.20)，对充分大的n可得

$$\begin{aligned}
\frac{x_{n+1}}{a_{n+1}^\delta} &= (a_n/a_{n+1})^\delta\left[\frac{1}{a_n^\delta}(x_n + a_n(Hx_n + D_n x_n) + a_n\varepsilon_{n+1}\right]\\
&= \left(1+\delta\,\frac{a_n - a_{n+1}}{a_{n+1}} + O\left(\left(\frac{a_n - a_{n+1}}{a_{n+1}}\right)^2\right)\right)\\
&\quad\times\left(\frac{x_n}{a_n^\delta} + a_n H\,\frac{x_n}{a_n^\delta} + a_n D_n\,\frac{x_n}{a_n^\delta} + a_n^{1-\delta}\varepsilon_{n+1}\right)\\
&= \frac{x_n}{a_n^\delta} + a_n\left(H + \delta\,\frac{a_n - a_{n+1}}{a_n a_{n+1}}I + c_n\right)\frac{x_n}{a_n^\delta}\\
&\quad + a_n\left(\frac{e_{n+1}}{a_n^\delta} + \frac{\nu_{n+1}}{a_n^\delta} + O\left(\frac{a_n - a_{n+1}}{a_{n+1}}\right)\frac{\varepsilon_{n+1}}{a_n^\delta}\right),
\end{aligned} \tag{3.2.23}$$

其中

$$c_n=\left(\delta\frac{a_n-a_{n+1}}{a_{n+1}}+O\left(\left(\frac{a_n-a_{n+1}}{a_{n+1}}\right)^2\right)+1\right)D_n$$
$$+O\left(\frac{a_n-a_{n+1}}{a_{n+1}}\right)^2\frac{1}{a_n}I\xrightarrow[n\to\infty]{}0.$$

而由式(3.2.4)和(3.2.5)知,

$$\frac{\nu_{n+1}}{a_n^\delta}\xrightarrow[n\to\infty]{}0,\ O\left(\frac{a_n-a_{n+1}}{a_{n+1}}\right)\frac{\varepsilon_{n+1}}{a_n^\delta}\xrightarrow[n\to\infty]{}0. \tag{3.2.24}$$

若记

$$H_n\triangleq H+\delta\frac{a_n-a_{n+1}}{a_na_{n+1}}I+c_n,\ y_n\triangleq\frac{x_n}{a_n^\delta},$$
$$e'_{n+1}\triangleq\frac{e_{n+1}}{a_n^\delta},\ \nu'_{n+1}=\frac{\nu_{n+1}}{a_n^\delta}+O\left(\frac{a_n-a_{n+1}}{a_{n+1}}\right)\frac{\varepsilon_{n+1}}{a_n^\delta}$$

则可重写式(3.2.23)为

$$y_{n+1}=y_n+a_nH_ny_n+a_n(e'_{n+1}+\nu'_{n+1}).$$

注意到 $H_n\xrightarrow[n\to\infty]{}H+\delta\alpha I$, 由式(3.2.2)知, $H+\delta\alpha I$ 稳定, 注意到条件 A 3.2.2, 便知引理 3.2.1 的条件成立, 所以 $y_n\xrightarrow[n\to\infty]{}0$. □

利用刚得到的定理 3.3.1 来看一下变界截尾的 RM 算法或 RM 算法的强收敛速度与步长因子$\{a_n\}$的关系. 为便于描述, 取 $a_n=\frac{1}{n^\alpha}$, $\frac{1}{2}<\alpha\leqslant1$, 取量测误差$\{\varepsilon_n\}$为如式(3.2.7)所描述的$\{e_n\}$. 这时, 为使 $\sum_{n=1}^{\infty}a_n^{1-\delta}e_{n+1}$ a.s. 收敛, 只要要求

$$\sum_{n=2}^{\infty}\mathrm{E}\left(\frac{1}{n^{2(\alpha-\alpha\delta)}}\|\varepsilon_n\|^2|\mathscr{F}_{n-1}\right)<\infty \tag{3.2.25}$$

或者 $2(\alpha-\alpha\delta)>1$, 即 $\delta<1-\frac{1}{2\alpha}$.

于是, 由 $\frac{1}{2}<\alpha\leqslant1$, 可得当 $\alpha=1$ 时, 算法有最好的收敛速度 $o\left(\frac{1}{n^\delta}\right)$, $\forall\delta<\frac{1}{2}$. 当 $\frac{1}{2}<\alpha<1$ 时, 算法的收敛速度为 $o\left(\frac{1}{n^{\alpha\delta}}\right)$, 由 $\alpha\delta<\alpha\left(1-\frac{1}{2\alpha}\right)=\alpha-\frac{1}{2}$, 因而当 α 趋于 $\frac{1}{2}$, 算法的收敛速度减慢,

直至当 $\alpha=\frac{1}{2}$，无法保证式(3.2.25)成立. 从这些分析可以看出，α 在 $\frac{1}{2}<\alpha\leqslant 1$ 范围内的大小，决定了 RM 算法的强收敛速度. 也就是说，步长因子$\{a_n\}$向零的衰减速度决定着算法的强收敛速度.

§3.3 一些数值模拟例子

本节所给出的数值模拟例子中，$\{\omega_n\}$是计算机模拟的独立同分布随机变量序列，近似地 $\omega_i\sim N(0,\ \sigma)$，量测误差序列$\{\varepsilon_n\}$的模型为形如

$$\varepsilon_n+A_1\varepsilon_{n-1}=\omega_n+C_1\omega_{n-1} \tag{3.3.1}$$

的 ARMA 序列.

例 3.3.1 一维单零点的线性回归函数.

取 $h(x)=-3(x-1), C_1=-2, A_1=0.2, \sigma=0.1, a_n=\frac{1}{1+n}$. 显然 $v(x)=(x-1)^2$ 可作为 Lyapunov 函数. 由于 $h(x)$ 是线性函数，采用通常 RM 算法，取 $x_0=10$.

计算结果：$x_{10}=1.06,\ x_{20}=1.02, x_{80}=1.01, x_i=1\pm 0.01(i>600)$，$x_{1300}=0.99908$.

例 3.3.2 一维单零点非线性回归函数.

$h(x)=-(x-10)^3, C_1=0.5,\ A_1=-0.9,\ \sigma=0.1, a_n=\frac{1}{1+n}$，$v(x)=(x-10)^2$ 可作为 Lyapunov 函数，所以变界截尾 RM 算法应该收敛. 取截尾界序列 $M_n=2^{n+1}$，截尾返回点 $x^*=0.5$.

计算结果：$x_{10}=0.5$，$x_{20}=43.4$，$x_{30}=-23.9$，$x_{40}=21.9$，$x_{50}=7.76$，$x_{60}=8.74$，$x_{70}=8.97$，$x_{100}=9.26$，$x_{200}=9.46$，$x_{300}=9.52$，$x_{400}=9.61$.

例 3.3.3 三维多零点线性回归函数.

$$h(x)=\begin{pmatrix}-1 & 1 & 2\\ 6 & -12 & -12\\ 1 & -1 & -2\end{pmatrix}x\triangleq Ax,$$ 其中 A 阵不满秩，具有非正

实部本征值. $C_1=\begin{pmatrix}2&0&0\\0&2&0\\0&0&2\end{pmatrix}$, $A_1=\begin{pmatrix}-0.5&0&0\\0&-0.5&0\\0&0&-0.5\end{pmatrix}$,
$\sigma=0.1$, $a_n=\frac{1}{1+n}$, 采用通常 RM 算法求零点, $x_0=(50, 100, -300)^T$.

计算结果:

$$x_{10}=\begin{pmatrix}3795\\-42364\\-4046\end{pmatrix},\ x_{20}=\begin{pmatrix}-168.0\\-1.59\\-82.6\end{pmatrix},\ x_{30}=\begin{pmatrix}-167.6\\-0.999\\-83.0\end{pmatrix},$$

$$x_{40}=\begin{pmatrix}-167.4\\-0.668\\-83.1\end{pmatrix},\ x_{50}=\begin{pmatrix}-167.3\\-0.478\\-83.2\end{pmatrix},\ x_{60}=\begin{pmatrix}-167.3\\-0.339\\-83.3\end{pmatrix},$$

$$x_{70}=\begin{pmatrix}-167.3\\-0.264\\-83.4\end{pmatrix}\approx\begin{pmatrix}-168\\0\\-84\end{pmatrix}\in J=\begin{pmatrix}2\alpha\\0\\\alpha\end{pmatrix},$$

α 为任意实数.

例 3.3.4 一维单极值非线性回归函数

$$h(x)=(x+10)^4,\ C_1=2,\ A_1=-0.5,\ \sigma=0.01,\ a_n=\frac{1}{1+n},$$

采用变界截尾 KW 算法求极值, 量测自变量差序列

$$b_n=\frac{1}{1+n},\ M_n=2n,\ x^*=0.5,\ v(x)=(x+10)^2.$$

计算结果: $x_{10}=0.5$, $x_i=-10\pm0.1$, $i\geqslant220$.

从以上数值模拟例子可以看出, 本章前面所得到的理论结果都得到验证. 但当算法序列越靠近真值, 算法收敛得越慢, 当回归函数的非线性越强, 算法也收敛得越慢, 这正是随机逼近算法的一个局限性. 因此, 如何改进算法的收敛速度, 使更适应实际需要, 是一个很有意义的问题. 本书将要讨论的渐近有效性和并行处理都和这一问题有关.

§3.4 算法收敛性对量测误差的一个充分必要条件

在本章的前面部分，我们已经对随机逼近算法强收敛的条件，无论是回归函数，还是量测误差，都得到了相当简明的结果. 但是，这些都还只是充分条件. 如果我们能获得收敛的充分必要条件，不仅理论上有意义，在实际应用中，还可以据此扩大适用范围.

为使算法收敛，至今还没有一般的必要条件. 但对量测误差已有一些结果. 1984年，在文献[27]中给出了RM算法关于量测误差的强收敛性充分必要条件. 他的结果是：如果 $a_n=\frac{1}{n}$，在回归函数满足与A 2.2.2和A 2.2.4相似的条件下，RM算法大范围强收敛的充分必要条件是量测误差应在算术平均意义下趋于0，也即

$$\frac{1}{n}\sum_{i=1}^{n}e_{i+1}\xrightarrow[n\to\infty]{}0,\quad \text{a.s.}. \tag{3.4.1}$$

由于这个结论中对回归函数的要求过强，在文献[13]中，运用变界截尾方法，将回归函数的限制显著地放宽. 例如，完全去掉了回归函数的线性增长条件. 在文献[12]中又将步长因子$\{a_n\}$取成较一般的形式.

下面，引入假设组A 3.4：

A 3.4.0 设回归函数$h(x)$是$R^l\to R^l$的连续函数，并有单零点x^0.

A 3.4.1 步长因子$\{a_n\}$满足$a_n>0$, $a_n\xrightarrow[n\to\infty]{}0$, $\sum_{n=1}^{\infty}a_n=\infty$，并存在$N_1$和$b>0$，当$n\geqslant N_1$时，

$$a_{n+1}\leqslant a_n\leqslant a_{n+1}(1+ba_n). \tag{3.4.2}$$

A 3.4.2 存在一定义在R^l上的二次连续可微实值函数$v(\cdot)$及实数$M>0$，使

$$x^*\in\{x:\|x\|<M\},\ d\triangleq\inf\{v(x):\|x\|=M\},\ v(x^*)<d, \tag{3.4.3}$$

$$v(x)\neq v(x^0),\ \forall x\neq x^0;\ v_{(x)}^{\tau}h(x)<0,\ \forall x\neq x^0, \qquad (3.4.4)$$

其中 x^* 由式(3.1.4)中给出.

A 3.4.3 量测量为

$$y_{n+1}=h(x_n)+\varepsilon_{n+1}, \qquad (3.4.5)$$

量测误差列$\{\varepsilon_n\}$满足

$$a_n\sum_{i=1}^{n}\varepsilon_{i+1}\xrightarrow[n\to\infty]{}0,\quad \text{a.s.}. \qquad (3.4.6)$$

在讨论必要条件时,对$\{a_n\}$还要用到下面的条件:

A 3.4.4 $\limsup\limits_{n\to\infty} na_n<\infty$, 并对充分大的 n,

$$1-\frac{\beta_1}{n+1}\leqslant\frac{a_{n+1}}{a_n}\leqslant 1-\frac{\alpha_1}{n+1},\ 0<\alpha_1<\beta_1<2. \qquad (3.4.7)$$

注 3.4.1 容易验证,下面一些$\{a_n\}$的例子满足 A 3.4.1.

i) $a_n=a/n$, 这时可取 $N_1\geqslant 1/(ba-1)$, $ba>1$.

ii) $a_n=a/\sqrt{n}$, 这时可取 $N_1\geqslant 1/(2b^2a^2)$.

iii) $a_n=a/\log n$, 这时可取 $N_1\geqslant 2/(ba)$.

同样可验证 $a_n=\dfrac{a}{n}$ 或 $a_n=\dfrac{a}{n\log n}$ 满足 A 3.4.4, 同时满足 A 3.4.1 和 A 3.4.4 的常见序列为 $\left\{\dfrac{a}{n}\right\}$.

引理 3.4.1 在假设 A 3.4.1 和 A 3.4.3 下,

$$a_n\varepsilon_{n+1}\xrightarrow[n\to\infty]{}0,\quad \text{a.s.}, \qquad (3.4.8)$$

并对任给的 $T>0$, $\delta>0$, 存在 $N>0$, 使得对任意 $n\geqslant N$, 和任意 m: $n\leqslant m\leqslant m(n,T)+1$, 都有

$$\left\|\sum_{i=n}^{m}a_i\varepsilon_{i+1}\right\|\leqslant\delta T, \qquad (3.4.9)$$

其中 $m(n,T)$的定义如式(2.4.3)所示.

证明 由式(3.4.2)和式(3.4.6)得

$$a_n\varepsilon_{n+1}=a_n\left(\sum_{i=1}^{n}\varepsilon_{i+1}-\sum_{i=1}^{n-1}\varepsilon_{i+1}\right)$$

$$=a_n\sum_{i=1}^{n}\varepsilon_{i+1}-\frac{a_n}{a_{n-1}}\cdot a_{n-1}\sum_{i=1}^{n-1}\varepsilon_{i+1}\xrightarrow[n\to\infty]{}0,\quad \text{a.s.},$$

用分部求和法知，存在 N，只要 $n\geqslant N$，对任 m：$n\leqslant m\leqslant m(n, T)+1$，都有

$$\left\|\sum_{i=n}^{m} a_i\varepsilon_{i+1}\right\|=\left\|a_m\sum_{i=1}^{m}\varepsilon_{i+1}-a_n\sum_{i=1}^{n-1}\varepsilon_{i+1}+\sum_{i=n}^{m-1}(a_i-a_{i+1})\sum_{j=1}^{i}\varepsilon_{j+1}\right\|$$

$$\leqslant\frac{1}{2}\delta T+b\sum_{i=n}^{m-1}a_{i+1}\left\|a_i\sum_{j=1}^{i}\varepsilon_j\right\|$$

$$\leqslant\frac{1}{2}\delta T+\frac{1}{2}\delta T=\delta T. \quad \square$$

现在，我们给出本节的主要定理.

定理 3.4.1 在假设 A 3.4.0、A 3.4.1 和 A 3.4.2 下，变界截尾的 RM 算法(3.4.5)、(3.1.3)和(3.1.4)对任意初值 x_0 都有

$$\lim_{n\to\infty} x_n=x^0, \quad \text{a.s.}, \tag{3.4.10}$$

的充分条件为式(3.4.6)成立.

而在条件 A 3.4.2 和 A 3.4.4 下，若上述算法中 $x^*\neq x^0$(或当 $\sigma(n)$为偶数时，取 $x^*=x_1^*$，当 $\sigma(n)$为奇数时，取 $x^*=x_2^*$，$x_1^*\neq x_2^*$)，则式(3.4.10)成立的必要条件是式(3.4.6)成立.

因此，在 A 3.4.0、A 3.4.2 的条件下，当 $a_n=\frac{a}{n}$ 时，为保证式(3.4.10)成立的充分必要条件是式(3.4.6)成立.

证明 注意到式(3.4.9)中的 δ 可任意小，所以式(2.4.4)成立. 那么式(3.4.10)直接得自定理3.1.1.

现在设 A 3.4.2 和 A 3.4.4 成立，同时上述算法 $x_n\xrightarrow[n\to\infty]{} x^0$，故由 $x^*\neq x^0$，算法中的截尾次数必定是有限的. 即存在 N，当 $i\geqslant N$ 时，算法成了通常的 RM 算法.

$$x_{i+1}=x_i+a_ih(x_i)+a_i\varepsilon_{i+1}, \quad i\geqslant N. \tag{3.4.11}$$

在上式两端同乘以 i，然后再将所得等式从 N 加到 n，则有

$$x_{n+1}=\frac{1}{n}(N-1)x_N+\frac{1}{n}\sum_{i=N}^{n}x_i+\frac{1}{n}\sum_{i=N}^{n}ia_ih(x_i)$$
$$+\frac{1}{n}\sum_{i=N}^{n}ia_i\varepsilon_{i+1}, \tag{3.4.12}$$

由于 $x_n\xrightarrow[n\to\infty]{} x^0$ a.s.，并且 $\limsup\limits_{n\to\infty} ia_i<\infty$，所以从式(3.4.12)可知，

$$\frac{1}{n}\sum_{i=N}^{n} i a_i \varepsilon_{i+1} \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.}, \tag{3.4.13}$$

而从 A 3.4.4 知，存在某常数 $\xi \in (0, 1)$，使得

$$1+\frac{\xi-2}{n+1} \leqslant \frac{a_{n+1}}{a_n} \leqslant 1-\frac{\xi}{n+1}, \tag{3.4.14}$$

进而有

$$\left|\frac{(n+1)a_{n+1}}{a_n}-n\right| \leqslant 1-\xi. \tag{3.4.15}$$

记

$$A_n = \sum_{i=1}^{n} \varepsilon_{i+1},$$

于是有

$$\begin{aligned}\frac{1}{n}\sum_{i=N}^{n} i a_i \varepsilon_{i+1} &= \frac{1}{n}\sum_{i=N}^{n} i a_i (A_{i+1}-A_i) \\ &= a_n A_{n+1} - \frac{N}{n} a_N A_N + \frac{1}{n}\sum_{i=N}^{n-1}(i a_i - (i+1) a_{i+1}) A_{i+1},\end{aligned} \tag{3.4.16}$$

再利用式(3.4.13)、(3.4.15)和(3.4.16)，可知

$$\begin{aligned}\|a_n A_{n+1}\| &\leqslant \left\|\frac{1}{n}\sum_{i=N}^{n} i a_i \varepsilon_{i+1}\right\| + \frac{1}{n}\sum_{i=N}^{n-1}\|a_i A_{i+1}\|(1-\xi) \\ &\quad + \frac{N}{n}\|a_N A_N\| \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.},\end{aligned}$$

注意到式(3.4.14)，由此便知式(3.4.6)成立. □

从定理必要条件的证明过程中，特别是从式(3.4.12)、(3.4.13)和(3.4.16)可见，这时的 a_n 与 $\frac{1}{n}$ 的衰减速度相近，因此，讨论必要条件时对$\{a_n\}$的限制，要比充分条件中的$\{a_n\}$强得多. 不过，由于在实际应用中，$\{a_n\}$的选择是人为的，而且在§3.3中我们已分析了当 $a_n=\frac{a}{n}$ 时算法具有很好的强收敛速度，所以不妨就取 $a_n=\frac{a}{n}$ 为步长因子，而这时由本定理知，量测误差只要且至少要算术平均为 0.

注意到 $a_n \xrightarrow[n\to\infty]{} 0$，如果再设 a_n 非增，用 Kronecker 引理，从 A 2.3.3 立即推出式(3.4.6).

最后需要指出,对较一般的$\{a_n\}$,例如对满足 A2.3.1 的$\{a_n\}$,式(3.4.6)并不是“$x_n \xrightarrow[n\to\infty]{} x^0$”的充分必要条件,文献[16]中已举出了反例.

定理 3.4.2 设 A2.4.0~A2.4.2 成立,且 $J=x^0$,$h(\cdot)$在x^0处连续,那么 $x_n\to x^0$ 的充分必要条件是 A2.4.3 成立.

证明:(略)

§3.5 算法的稳健性分析

当算法收敛性条件不能严格满足时,算法会有怎样的渐近表现,这就是稳健性(robustness)分析.在实际应用中,人们都希望所用的算法的收敛条件能經得起小的扰动,也就是说,收敛条件的一些小偏差,不致造成算法的巨大偏差,甚至发散,也就是希望算法具有稳健性.稳健性研究工作参见文献[12,13]

仍设量测为

$$y_{n+1}=h(x_n)+\varepsilon_{n+1},\quad n\geqslant 1, \tag{3.5.1}$$

x_n 是对 x^0 的第 n 次估计,ε_{n+1} 是量满误差.

我们来研究变界截尾的 RM 算法(3.1.3)、(3.1.4)的稳健性.我们要用下面的条件.

A3.5.0 $h(\cdot)$是 $R^l\to R^l$ 的未知连续函数,$x^0\in R^l$,(我们不要求 $h(x^0)=0$).

A3.5.1 步长因子 a_n 非增,$a_n\to 0$,$\sum_{n=1}^{\infty}a_n=\infty$,并存在 N_0,对 $n\geqslant N_0$,

$$\frac{1}{a_{n+1}}-\frac{1}{a_n}\leqslant 1,\text{即 } a_n-a_{n+1}\leqslant a_na_{n+1}. \tag{3.5.2}$$

A3.5.2 存在二次连续可微 $R^l\to R$ 的非负函数 $v(\cdot)$:$v(x^0)=0$,$\lim\limits_{\|x\|\to\infty}v(x)=\infty$,并且

$$v_x^{\tau}(x)h(x)<0,\ \forall x:\ \|x-x^0\|>e\geqslant 0 \tag{3.5.3}$$

A3.5.3 量测误差$\{\varepsilon_n\}$满足

$$\limsup_{n\to\infty}\left\|a_n\sum_{i=1}^{n}e_{i+1}\right\|\leqslant e<\infty,\tag{3.5.4}$$

e 可依赖于 ω(样本).

和上一节的条件相比，这里的 ε 和 e 不一定是 0. 这表示使算法收敛的条件不严格满足，我们要研究算法给出的估计对 x^0 的渐近偏差如何依赖于 ε 和 e.

作为算法稳健分析的第一步是讨论算法的有界性. 这是最粗略的稳健性，并且一旦算法有界，当 n 充分大时，截尾算法就成了非截尾的算法，处理较简单. 另一方面，当使算法收敛的条件不严格满足时，对较一般的回归函数，目前还不知道通常的 RM 算法是否还能有界.

在这一节下面的讨论中，样本都是固定的.

在算法(3.1.3)、(3.1.4)中我们用了一个固定点 x^*，当算法超出一个随时间变化的活动界限时，就把算法拉回到 x^*. 记

$$d_0=v(x^*).\tag{3.5.5}$$

由于 $v(x)\xrightarrow[\|x\|\to\infty]{}\infty$，所以可取 M 充分大，使

$$\|x^*\|<M,\quad \inf\{v(x):\|x\|>M\}\triangleq d>d_0.\tag{3.5.6}$$

取算法(3.1.3)、(3.1.4)中的初始截尾界 M_0 适当大，使

$$M_0>M+8,\tag{3.5.7}$$

取实数 $0<\delta_1<\delta_2$，使

$$[\delta_1,\ \delta_2]\subset(d_0,\ d),\quad \frac{d-d_0}{2}<\delta_1,\tag{3.5.8}$$

由于 $v(\cdot)$连续，$v(x^0)=0$，所以存在 $e^*>0$，使

$$\{x:v(x)\geqslant\delta_1\}\subset\{x:\|x-x^0\|>e^*\},\tag{3.5.9}$$

记

$$D=\{x:\delta_1\leqslant v(x)\leqslant\delta_2\},\tag{3.5.10}$$

$$h_1=\max_{\|x\|\leqslant M+8}\|h(x)\|,\tag{3.5.11}$$

$$U=\{x:\|x\|\leqslant M+6\},\tag{3.5.12}$$

$$r_1=\max_{x\in U}\|v_x(x)\|,\quad r_2=\max_{x\in U}\|v_{xx}(x)\|,\tag{3.5.13}$$

这里 $v_{xx}(x)$表示由 $v(x)$的二阶偏导数所构成的阵.

由于 $\delta_2<d$，所以对 D 中的任一 x 必有 $\|x\|\leqslant M$. 因此 $D\subset U$，还记

$$a=-\max_{x\in D} h^{\tau}(x)v_x(x). \tag{3.5.14}$$

下面只讨论式(3.5.3)中的 $\varepsilon\leqslant\varepsilon^*$ 的情形. 从式(3.5.9)、(3.5.10)知，对 $x\in D$，必有 $\|x-x^0\|>\varepsilon^*\geqslant\varepsilon$. 因此，由式(3.5.14)定义的 $a>0$.

由于 $h(\cdot)$ 及 $v(\cdot)$ 连续，则存在 $\alpha>0$，$\beta>0$，$\varepsilon^*\in(0, T]$，T^* 使下面诸不等式成立：

$$T^*\leqslant\frac{1}{1+h_1} \tag{3.5.15}$$

$$-\alpha>-a+\frac{1}{2}r_2(2+h_1)^2T^*, \tag{3.5.16}$$

$$-\beta>-\alpha+r_1\cdot\max_{\substack{x,y\in U\\ \|x-y\|\leqslant 3\varepsilon^*+T^*(2+h_1)}}\|h(x)-h(y)\|, \tag{3.5.17}$$

$$\max_{\substack{\|x-y\|\leqslant 4\varepsilon^*\\ x,y\in U}}|v(x)-v(y)|<\delta_2-\delta_1, \tag{3.5.18}$$

$$\varepsilon^*<\beta T^*/(8r_1+14r_2), \tag{3.5.19}$$

$$\delta_1+r_1[7\varepsilon^*+(h_1+2\varepsilon^*)T^*]<\delta_2. \tag{3.5.20}$$

当 $\varepsilon\leqslant\varepsilon^*$ 时，用 A 3.5.3，可取充分大的 $N\geqslant N_0$，使对 $n\geqslant N$ 成立

$$a_n\leqslant\varepsilon^*/(2h_1),\ a_n\left\|\sum_{i=1}^{n}\varepsilon_{i+1}\right\|\leqslant\frac{11}{8}\varepsilon^*. \tag{3.5.21}$$

引理 3.5.1 设 A 3.5.0, A 3.5.1, A 3.5.3 成立，并且式(3.5.4)中的 $\varepsilon\leqslant\varepsilon^*\leqslant 1$，算法(3.1.3)、(3.1.4)中的 M_0 满足式(3.5.7)，如果对某个 $n(\geqslant N)$ 成立 $\|x_n\|\leqslant M$，那么对任意 $T\in\left[0, \frac{1}{1+h_1}\right]$ 有

$$\left\|\sum_{i=n}^{m+1}a_iy_{i+1}\right\|\leqslant 6\varepsilon^*+2,\ \forall m: n\leqslant m\leqslant m(n, T). \tag{3.5.22}$$

证明 由于当 T 上升时，$m(n, T)$ 非降，所以引理只要对 $T=\frac{1}{1+h_1}$ 证明.

反证法. 假设引理不成立，那么存在 m_1，使

$$m_1=\min\left\{m: n\leqslant m\leqslant m(n, T), \left\|\sum_{i=n}^{m+1} a_i y_{i+1}\right\|>6\varepsilon^*+2\right\}.$$

那么对任意 $m: n\leqslant m\leqslant m_1$ 就有

$$\left\|x_n+\sum_{i=n}^{m} a_i y_{i+1}\right\|\leqslant\|x_n\|+\left\|\sum_{i=n}^{m} a_i y_{i+1}\right\|$$
$$\leqslant M+6\varepsilon^*+2<M_0<M_{\sigma_n}, \tag{3.5.23}$$

所以从式(3.1.3)知:

$$x_{m+1}=x_m+a_m y_{m+1},\ \forall m: n\leqslant m\leqslant m_1, \tag{3.5.24}$$

由此从 m_1 的定义知

$$\|x_{m_1+1}-x_n+a_{m_1+1}y_{m_1+2}\|=\left\|\sum_{i=n}^{m_1+1} a_i y_{i+1}\right\|>6\varepsilon^*+2. \tag{3.5.25}$$

另一方面, 从式(3.5.23)、(3.5.24)知

$$\|x_m\|\leqslant M+6\varepsilon^*+2,\ \forall m: n\leqslant m\leqslant m_1+1, \tag{3.5.26}$$

又从式(3.5.11)知

$$\|h(x_m)\|\leqslant h_1,\ \forall m: n\leqslant m\leqslant m_1+1. \tag{3.5.27}$$

注意到式(3.5.2)和(3.5.21)便知

$$\left\|\sum_{i=n}^{m_1} a_i\varepsilon_{i+1}\right\|=\left\|a_{m_1}\sum_{i=1}^{m_1}\varepsilon_{i+1}-a_n\sum_{i=1}^{n-1}\varepsilon_{i+1}+\sum_{i=n}^{m_1-1}(a_i-a_{i+1})\sum_{j=1}^{i}\varepsilon_{j+1}\right\|$$
$$\leqslant\frac{11}{4}\varepsilon^*+\sum_{i=n}^{m_1-1}a_{i+1}\left\|a_i\sum_{j=1}^{i}\varepsilon_{j+1}\right\|$$
$$\leqslant\frac{11}{4}\varepsilon^*+\frac{11}{8}T\varepsilon^*. \tag{3.5.28}$$

用式(3.5.27)、(3.5.28)就得到

$$\|x_{m_1+1}-x_n+a_{m_1+1}y_{m_1+2}\|\leqslant\|x_{m_1+1}-x_n\|$$
$$+\|a_{m_1+1}(h(x_{m_1+1})+\varepsilon_{m_1+2})\|\leqslant\left\|\sum_{i=n}^{m_1}a_i h(x_i)\right\|$$
$$+\left\|\sum_{i=n}^{m_1}a_i\varepsilon_{i+1}\right\|+a_{m_1+1}h_1+a_{m_1+1}\left\|\sum_{i=1}^{m_1+1}\varepsilon_{i+1}-\sum_{i=1}^{m_1}\varepsilon_{i+1}\right\|$$
$$\leqslant Th_1+\frac{11}{4}\varepsilon^*+\frac{11}{8}T\varepsilon^*+\frac{\varepsilon^*}{2}+\frac{11}{4}\varepsilon^*$$
$$=\left(h_1+\frac{11}{8}\varepsilon^*\right)T+6\varepsilon^*<2+6\varepsilon^*.$$

这和式(3.5.25)矛盾, 也就证明引理. □

引理3.5.2 在引理3.5.1的条件，对任意 $T\in\left[0, \frac{1}{1+h_1}\right]$ 成立

$$\|x_m-x_n\|<3\varepsilon^*+T(h_1+2\varepsilon^*),\ \forall m: n\leqslant m\leqslant m(n, T)+1. \tag{3.5.29}$$

证明 由于 $\|x_n\|\leqslant M$，用引理3.5.1知

$$\left\|x_n+\sum_{i=n}^{m}a_i y_{i+1}\right\|\leqslant M+6\varepsilon^*+2\leqslant M_0\leqslant M_{\sigma_n},$$

$$\forall m: n\leqslant m\leqslant m(n, T)+1.$$

根据算法的定义知

$$x_{m+1}=x_m+a_m y_{m+1},\quad \forall m: n\leqslant m\leqslant m(n, T)+1,$$

并且 $\|x_m\|\leqslant M+6\varepsilon^*+2,\quad \|h(x_m)\|\leqslant h_1.$

所以，

$$\begin{aligned}\|x_m-x_n\|\leqslant&\left\|\sum_{i=n}^{m-1}a_i(h(x_i)+\varepsilon_{i+1})\right\|\leqslant h_1T\\&+\left\|a_{m-1}\sum_{i=1}^{m-1}\varepsilon_{i+1}-a_n\sum_{i=1}^{n-1}\varepsilon_{i+1}+\sum_{i=n}^{m-2}(a_i-a_{i+1})\sum_{j=1}^{i}\varepsilon_{j+1}\right\|\\\leqslant&\ h_1T+\frac{11}{4}\varepsilon^*+\frac{11}{8}T\varepsilon^*<3\varepsilon^*+T(h_1+2\varepsilon^*),\\&\forall m: n\leqslant m\leqslant m(n, T)+1.\end{aligned}$$ □

当收敛条件不严格满足时，也就是当 $\theta>0$, $\varepsilon>0$ 时，下面来证式(3.1.3)、(3.1.4)给出的 x_n 仍对 n 一致有界.

定理3.5.1 设A3.5.0～A3.5.3成立，θ^* 满足式(3.5.9)，$\varepsilon^*\leqslant1$. 那么对固定的样本，存在不依赖于 $\varepsilon\in[0, \varepsilon^*]$ 及 $\theta\in[0, \theta^*]$ 的 σ，使

$$\sup_n\sigma_n\leqslant\sigma<\infty,$$

即由式(3.1.3)、(3.1.4)给出的 x_n 对 n 一致有界.

证明 反证法. 假设定理不成立，设正整数 r_0 充分大，使

$$a_{r_0}<T^*,\ r_0>N, \tag{3.5.30}$$

这里 $N\geqslant N_0$，并且当 $n\geqslant N$ 时式(3.5.21)成立.

从定理的反假设知，存在 $\varepsilon\in[0, \varepsilon^*]$ 及 n 使 $\sigma_n>r_1\triangleq r_0+2$.

设 n_0 是使下式成立的最大时间:

$$r_0=\sum_{i=1}^{n_0-1} I_{[\|x_i+a_i y_{i+1}\|>M_{\sigma_i}]}. \tag{3.5.31}$$

从式(3.1.4)及 n_0 的极大性知

$$\sigma_{n_0}=r_0,\ \sigma_{n_0+1}=r_0+1,\ x_{n_0+1}=x^*, \tag{3.5.32}$$

又从式(3.5.30)及(3.5.31)知

$$n_0>r_0>N. \tag{3.5.33}$$

我们先证在反假设条件下,必存在 $m_0\geqslant n_0+2$,使

$$\|x_{m_0}\|>M. \tag{3.5.34}$$

如果 $\|x_n\|\leqslant M,\ \forall n\geqslant n_0+2$, 那么从式(3.5.27)知

$$\begin{aligned}\|x_n+a_r y_{n+1}\| &\leqslant M+\|a_n h(x_n)\|+\|a_n e_{n+1}\| \\ &\leqslant M+a_n h_1+\left\|a_n\sum_{i=1}^{n}\varepsilon_{i+1}-a_n\sum_{i=1}^{n-1}\varepsilon_{i+1}\right\| \\ &\leqslant M+\frac{\varepsilon^*}{2}+\frac{11}{4}\varepsilon^*\leqslant M+4<M_0\leqslant M_{\sigma_n}, \\ &\qquad\qquad\forall n\geqslant n_0+2.\end{aligned} \tag{3.5.35}$$

这和式(3.5.32)一起表明

$$\sigma_n\equiv\sigma_{n_0+2}\leqslant 1+\sigma_{n_0+1}=r_0+2=r_1,\quad \forall n\geqslant n_0+2,$$

而这和反假设矛盾,所以式(3.5.34)成立.

从式(3.5.5)、(3.5.8)及(3.5.32)知

$$v(x_{n_0+1})=v(x^*)<\delta_1, \tag{3.5.36}$$

又从式(3.5.34)、(3.5.6)及(3.5.8)知

$$v(x_{m_0})\geqslant d>\delta_2,\quad m_0\geqslant n_0+2. \tag{3.5.37}$$

我们来证 $m_0>n_0+2$. 为此只要证明

$v(x_{n_0+2})<\delta_2$, 那么由式(3.5.37)说明 $m_0>n_0+2$,

注意到 $\|x_{n_0+1}\|=\|x^*\|<M$, 和式(3.5.35)类似地证明

$$\|x_{n_0+1}+a_{n_0+1}y_{n_0+2}\|\leqslant M+4<M_0\leqslant M_{\sigma_n}, \tag{3.5.38}$$

所以从式(3.1.3)知

$$x_{n_0+2}=x_{n_0+1}+a_{n_0+1}y_{n_0+2},\ \|x_{n_0+2}\|\leqslant M+4. \tag{3.5.39}$$

从式(3.5.35)及(3.5.39)还看出

$$\|x_{n_0+2}-x_{n_0+1}\|\leqslant\frac{13}{4}\varepsilon^*<4\varepsilon^*. \tag{3.5.40}$$

注意到 $x_{n_0+1}\in U$, $x_{n_0+2}\in U$, 从式(3.5.18)知

$$v(x_{n_0+2})<v(x_{n_0+1})+\delta_2-\delta_1.$$

用式(3.5.36), 知 $v(x_{n_0+2})<\delta_2$. 所以式(3.5.37)成为

$$v(x_{m_0})>\delta_2,\ m_0>n_0+2. \tag{3.5.41}$$

于是可以定义

$$n_1=\max\{i:\ v(x_{i-1})<\delta_1,\ n_0+2\leqslant i<m_0\},$$
$$n_2=\min\{i:\ v(x_i)>\delta_2,\ n_0+2<i\leqslant m_0\}.$$

这样就有

$$v(x_{n_1-1})<\delta_1,\ v(x_{n_2})>\delta_2, \tag{3.5.42}$$

$$\delta_1\leqslant v(x_i)\leqslant\delta_2,\ \forall i:\ n_1\leqslant i\leqslant n_2-1. \tag{3.5.43}$$

在引理 3.5.1、3.5.2 中取 $n=n_1$, $T=T^*$, 注意到式(3.5.6)以及 $v(x_{n_1})\leqslant\delta_2<d$, 便知 $\|x_{n_1}\|\leqslant M$. 用引理 3.5.1 及 3.5.2, 从式(3.5.30) 知 $a_{r_0}<T^*$. 注意到式(3.5.33) 及 $n_1\geqslant n_0+2>r_0$ 便知 $a_{n_1}<T^*$, 所以 $m(n, T^*)$确有定义.

用 Taylor 展式得

$$v(x_{m(n_1,T^*)+1})-v(x_{n_1})=(x_{m(n_1,T^*)+1}-x_{n_1})^\tau v_x(x_{n_1})$$
$$+\frac{1}{2}(x^\tau_{m(n_1,T^*)+1}-x^\tau_{n_1})v_{xx}(\eta)(x_{m(n_1,T^*)+1}-x_{n_1}), \tag{3.5.44}$$

这里 $\eta\in R^l$, 它的分量在 x_{n_1} 及 $x_{m(n_1,T^*)+1}$ 的相应分量之间.

现证 $v(x_{m(n_1,T^*)+1})<\delta_1$, 然后证明这个不等式将导致矛盾.

用引理 3.5.2 知

$$\|x_{m(n_1,T^*)+1}-x_{n_1}\|<3\varepsilon^*+T^*(h_1+2\varepsilon^*), \tag{3.5.45}$$
$$\|\eta-x_{n_1}\|\leqslant3\varepsilon^*+T^*(h_1+2\varepsilon^*),$$

所以

$$\|\eta\|\leqslant M+3\varepsilon^*+T^*(h_1+2\varepsilon^*)\leqslant M+3\varepsilon^*+\frac{h_1+2\varepsilon^*}{h_1+1}$$
$$\leqslant M+5\varepsilon^*+1\leqslant M+6.$$

从式(3.5.12)知 $\eta\in U$, 而从式(3.5.13)有 $\|v_{xx}(\eta)\|\leqslant r_2$. 用引理 3.5.1, 我们继续估计式(3.5.44):

$$v(x_{m(n_1,T^*)+1})-v(x_{n_1})\leqslant\sum_{i=n_1}^{m(n_1,T^*)}a_i(h^\tau(x_i)-\varepsilon_{i+1}^\tau)v_x(x_{n_1})$$

$$+\frac{r_2}{2}(3\varepsilon^*+T^*(h_1+2\varepsilon^*))^2$$

$$\leqslant\sum_{i=n_1}^{m(n_1,T^*)}a_ih^\tau(x_{n_1})x_x(x_{n_1})+\sum_{i=n_1}^{m(n_1,T^*)}a_i(h^\tau(x_i)-h^\tau(x_{n_1}))v_x(x_{n_1})$$

$$+\sum_{i=n_1}^{m(n_1,T^*)}a_i\varepsilon_{i+1}^\tau v_x(x_{n_1})+\frac{r_2}{2}(3\varepsilon^*+T^*(h_1+2\varepsilon^*))^2. \quad (3.5.46)$$

注意到 $v(x_{n_1-1})<\delta_1<\delta_2$，所以 $\|x_{n_1-1}\|\leqslant M$，$x_{n_1-1}\in U$. 把式(3.5.38)、(3.5.40)中的 n_0+1 换成 n_1-1 后，它们照样成立. 因此同(3.5.40)类似地得到

$$x_{n_1}\in U,\ \|x_{n_1}-x_{n_1-1}\|\leqslant\frac{13}{4}\varepsilon^*.$$

用式(3.5.13)及 Taylor 展开式便知

$$|v(x_{n_1})-v(x_{n_1-1})|\leqslant\|v_x(\theta)\|\cdot\|x_{n_1}-x_{n_1-1}\|$$

$$\leqslant\frac{13}{4}r_1\varepsilon^*,$$

所以

$$v(x_{n_1})\leqslant v(x_{n_1-1})+\frac{13}{4}r_1\varepsilon^*<\delta_1+\frac{13}{4}r_1\varepsilon^* \quad (3.5.47)$$

从式(3.5.43)知 $x_{n_1}\in D$. 把式(3.5.47)代入式(3.5.46)，并用式(3.5.14)就有

$$v(x_{m(n_1,T^*)+1})\leqslant\delta_1+\frac{13}{4}r_1\varepsilon^*-aT^*$$

$$+a_{m(n_1,T^*)+1}|h^\tau(x_{n_1})v_x(x_{n_1})|+r_1T^*\max_{n_1\leqslant i\leqslant m(n_1,T^*)}\|h(x_{n_1})$$

$$-h(x_i)\|+r_1\left\|\sum_{i=n_1}^{m(n_1,T^*)}a_i\varepsilon_{i+1}^\tau\right\|$$

$$+\frac{r_2}{2}[3\varepsilon^*+T^*(h_1+2\varepsilon^*)]^2.$$

对 $\sum_{i=n_1}^{m(n_1,T^*)}a_i\varepsilon_{i+1}$ 用类似于式(3.5.28)的估计知

$$v(x_{m(n_1,T^*)+1})\leqslant\delta_1+\frac{13}{4}r_1\varepsilon^*-aT^*+a_{n_1}h_1r_1$$

$$+r_1T^*\max_{n_1\leqslant i\leqslant m(n_1,T^*)}\|h(x_{n_1})-h(x_i)\|$$

$$+\left(\frac{11}{4}\varepsilon^*+\frac{11}{8}T^*\varepsilon^*\right)r_1+\frac{r_2}{2}[3\varepsilon^*+T^*(h_1+2\varepsilon^*)]^2. \tag{3.5.48}$$

注意到$\|x_{n_1}\|\leqslant M$，并用引理3.5.2知

$$\|x_m-x_{n_1}\|<3\varepsilon^*+T^*(h_1+2\varepsilon^*)\leqslant 3\varepsilon^*+T^*(h_1+2), \tag{3.5.49}$$

以及

$$\|x_m\|<M+3\varepsilon^*+T^*(h_1+2\varepsilon^*)<M+5\varepsilon^*+1\leqslant M+6,$$
$$\forall m: n_1\leqslant m\leqslant m(n_1,T^*)+1,$$

所以$x_m\in U$. 用式(3.5.17)，从式(3.5.48)得

$$v(x_{m(n_1,T^*)+1})\leqslant\delta_1+\frac{13}{4}r_1\varepsilon^*-\alpha T^*$$
$$+a_{n_1}h_1r_1+T^*(\alpha-\beta)+r_1\varepsilon^*\left(\frac{11}{4}+\frac{11}{8}T^*\right)$$
$$+\frac{r_2}{2}[3\varepsilon^*+T^*(h_1+2\varepsilon^*)]^2,$$

再用式(3.5.16)知

$$v(x_{m(n_1,T^*)+1})\leqslant\delta_1+\frac{13}{4}r_1\varepsilon^*$$
$$-T^*\left[\alpha+\frac{1}{2}r_2(h+2)^2T^*\right]+a_{n_1}h_1r_1$$
$$+T^*(\alpha-\beta)+r_1\varepsilon^*\left(\frac{11}{4}+\frac{11}{8}T^*\right)$$
$$+\frac{9}{2}r_2\varepsilon^{*2}+3r_2\varepsilon^*T^*(h_1+2\varepsilon^*)$$
$$+\frac{r_2}{2}T^{*2}(h_1+2\varepsilon^*)^2\leqslant\delta_1-\beta T^*$$
$$+r_1\varepsilon^*\left(6+\frac{11}{8}T^*\right)+a_{n_1}h_1r_1$$
$$+r_2\varepsilon^*\left(\frac{9}{2}\varepsilon^*+3T^*(h+2\varepsilon^*)\right).$$

从式(3.5.21)知 $a_{n_1}h_1 \leqslant \frac{\varepsilon^*}{2}$, 注意到式(3.5.15), 并用式(3.5.19), 得

$$v(x_{m(n_1, T^*)+1}) \leqslant \delta_1 - \beta T^* + 3r_1\varepsilon^* + 14r_2\varepsilon^* < \delta_1 \quad (3.5.50)$$

另一方面, 用引理 3.5.2 及式(3.5.13)、(3.5.20)及(3.5.47)知对任意 m: $n_1 \leqslant m \leqslant m(n_1, T^*)+1$.

$$\begin{aligned} v(x_m) &= v(x_{n_1}) + v_x^\tau(\phi)(x_m - x_{n_1}) \\ &\leqslant \delta_1 + \frac{13}{4} r_1\varepsilon^* + r_1[3\varepsilon^* + T^*(h_1 + 2\varepsilon^*)] < \delta_2, \end{aligned} \quad (3.5.51)$$

在上面 Taylor 展开式中 $\phi \in U$.

从式(3.5.42)、(3.5.43)及(3.5.51)知

$$m(n_1, T^*) + 1 < n_2,$$

不以 $v(x_{m(n_1, T^*)+1}) \geqslant \delta_1$, 这和式(3.5.50)矛盾, 说明反证法的假设所成立. □

推论 3.5.1 从定理 3.5.1 知存在 $\varepsilon^* > 0$, $e^* > 0$, 当 $e \in [0, {}^*e]$, $\varepsilon \in [0, \varepsilon^*]$时, 存在不依赖于 e 和 ε 的 σ, 使

$$\|x_n\| \leqslant M_\sigma, \quad (3.5.52)$$

并对充分大的 $N_1 (\geqslant N)$, 当 $n \geqslant N_1$ 时,

$$x_{n+1} = x_n + a_n(h(x_n) + \varepsilon_{n+1}), \quad (3.5.53)$$

也就是说, 算法(3.1.3)、(3.1.4)当 $n \geqslant N_1$ 时, 成了通常的 RM 算法.

记

$$\max_{\|x\| \leqslant M_\sigma + 8} v(x) = L_0, \quad \max_{\|x\| \leqslant M_\sigma} \|h(x)\| = L_3, \quad (3.5.54)$$

$$\max_{\|x\| \leqslant M_\sigma + 8} \|v_x(x)\| = L_1, \quad \max_{\|x\| \leqslant M_\sigma + 8} \|v_{xx}(x)\| = L_2. \quad (3.5.55)$$

任取 $\rho > e$, 记

$$\alpha(\rho) = \frac{1}{L_0} \min_{\substack{\|x - x^0\| \geqslant \rho \\ \|x\| \leqslant M_\sigma}} -h^\tau(x) v_x(x). \quad (3.5.56)$$

从式(3.5.3)知, $\alpha(\rho) > 0$. 再记

$$\beta(\rho)=\max_{\|x-x^0\|\leqslant\rho} h^{\tau}(x)v_a(x). \qquad (3.5.57)$$

若式(3.5.3)中的 $e=0$，则 $\beta(\rho)\leqslant 0$，否则它可能大于 0.

现在给出算法的稳健性结果.

定理 3.5.2 设 A 3.5.0～A 3.5.3 成立，x_n 由式(3.1.3)、(3.1.4)给出，M_0 满足式(3.5.7). 那么存在 $0<\varepsilon^*\leqslant 1$，$e^*>0$，以及定义在$[0,\infty)$上的非降、左连续函数 $f(t)\xrightarrow[t\to 0^+]{} 0$，只要 A 3.5.2 中的 $e\leqslant e^*$，A 3.5.3 中的 $\varepsilon\leqslant\varepsilon^*$，就有

$$\limsup_{n\to\infty}\|x_n-x^0\|$$

$$\leqslant f\left(\frac{3}{\alpha(\rho)}(\beta(\rho)\vee 0)+\frac{3L_4}{\alpha(\rho)}\varepsilon+3L_1\rho+24L_1\varepsilon\right)+8\varepsilon,$$

$$\forall\rho>e. \qquad (3.5.58)$$

实际上 $f(t)$ 可取为 $m(r)\triangleq\min\limits_{\|x-x^0\|\geqslant r} v(x)$ 的逆函数：$f(t)=\min\{r: m(r)=t\}$.

推论 3.5.2 当式 (3.5.3) 中的 $e=0$ 时 ($h(x^0)$ 不一定等于 0)，则$\beta(\rho)\leqslant 0$，那么式(3.5.58)右端为

$$f\left(\frac{3L_4}{\alpha(\rho)}\varepsilon+3L_1\rho+24L_1\varepsilon\right)+8\varepsilon\xrightarrow[\varepsilon\to 0]{} f(3L_1\rho).$$

因为 $e=0$，所以 ρ 可取任意小，所以$f(3L_1\rho)$可以任意小，也就是说估计误差$\|x_n-x^0\|$可以任意小. 这时当 A 3.5.3 中 $\varepsilon=0$ 时，那么在式(3.5.58)中两边先把 $\varepsilon\to 0$，再把 $\rho\to 0$ 就得

$$\limsup_{n\to\infty}\|x_n-x^0\|=0,$$

而当 $e>0$ 时，ρ 必须大于 e. 当 $\varepsilon\to 0$ 时，式(3.5.58)右端趋于

$$f\left(\frac{3}{\alpha(\rho)}(\beta(\rho)\vee 0)+3L_1\rho\right),$$

所以当 $\varepsilon\to 0$ 时，估计误差的大小取决于 $\dfrac{3}{\alpha(\rho)}(\beta(\rho)\vee 0)+3L_1\rho$ 的大小. 如果式(3.5.3)中的 $e\to 0$，则 $\rho\to 0$，这时估计误差取决于 $\dfrac{3}{\alpha(\rho)}(\beta(\rho)\vee 0)$的大小.

定理 3.5.2 的证明 给定 u_0 后，递推地定义

$$u_{n+1}=u_n+a_n(-u_n+e_{n+1}). \tag{3.5.59}$$

我们来证，只要 $N_2(\geqslant N_1)$ 充分大，就有

$$\sup_{n\geqslant N_2}\|u_n\|\leqslant 8\varepsilon. \tag{3.5.60}$$

记 $s_i=\sum_{j=1}^{i}e_{j+1}$，并认为 $\prod_{j=n+k}^{n}(1-a_j)=1,\ \forall k\geqslant 1$. 从式(3.5.59)知

$$\begin{aligned}\|u_{n+1}\|&=\left\|\prod_{i=N_1}^{n}(1-a_i)u_{N_1}+\sum_{i=N_1}^{n}\prod_{j=i+1}^{n}(1-a_j)a_ie_{i+1}\right\|\\&\leqslant\|u_{N_1}\|\exp\left[-\sum_{i=N_1}^{n}a_i\right]+\left\|\sum_{i=N_1}^{n}\prod_{j=i+1}^{n}(1-a_j)a_i(s_i-s_{i-1})\right\|.\end{aligned}$$

我们认为 N_1 已充分大，根据 A 3.5.3,

$$\|a_ns_n\|\leqslant 2\varepsilon,\ \forall n\geqslant N_1-1 \tag{3.5.61}$$

那么从上面得

$$\begin{aligned}\|u_{n+1}\|&\leqslant o(1)+\left\|a_ns_n-\prod_{j=N_1}^{n}(1-a_j)a_{N_1-1}s_{N_1-1}\right.\\&\quad\left.+\sum_{i=N_1-1}^{n-1}(1-a_j)a_is_i-\sum_{i=N_1-1}^{n-1}\prod_{j=i+2}^{n}(1-a_j)a_{i+1}s_i\right\|\\&\leqslant o(1)+4\varepsilon+\sum_{i=N_1-1}^{n-1}\prod_{j=i+2}^{n}\|(1-a_j)[a_i(1-a_{i+1})-a_{i+1}]s_i\|\\&\leqslant o(1)+4\varepsilon+2\varepsilon\sum_{i=N_1-1}^{n-1}\left|1-a_{i+1}-\frac{a_{i+1}}{a_i}\right|\exp\left(-\sum_{j=i+2}^{n}a_j\right).\end{aligned} \tag{3.5.62}$$

因为 a_n 非增，所以由式(3.5.2)知

$$-a_{n+1}\leqslant 1-\frac{a_{n+1}}{a_n}-a_{n+1}\leqslant 0. \tag{3.5.63}$$

不失一般性，可认为 $a_n<\frac{1}{2},\ \forall n\geqslant N_1$，所以由式(3.5.2)知

$$\frac{a_n}{a_{n+1}}\leqslant 1+a_n<\frac{3}{2}. \tag{3.5.64}$$

把式(3.5.63)、(3.5.64)用到式(3.5.62)中，得

$$\|u_{n+1}\|\leqslant o(1)+4\varepsilon+2\varepsilon\exp\left(-\sum_{j=1}^{n}a_j\right)\sum_{i=N_1-1}^{n-1}a_{i+1}\exp\left(\sum_{j=1}^{i+1}a_j\right)$$

$$\leqslant o(1)+4\varepsilon+3\varepsilon\exp\Big(-\sum_{j=1}^{n}a_j\Big)\sum_{i=N_1-1}^{n-1}a_{i+2}\exp\Big(\sum_{j=1}^{i+1}a_j\Big)$$
$$\leqslant o(1)+4\varepsilon+3\varepsilon\exp\Big(-\sum_{j=1}^{n}a_j\Big)\sum_{i=N_1-1}^{n-1}(e^{a_{i+2}}-1)\exp\Big(\sum_{j=1}^{i+1}a_j\Big)$$
$$=o(1)+4\varepsilon+3\varepsilon\exp\Big(-\sum_{j=1}^{n}a_j\Big)\Big(\exp\Big(\sum_{j=1}^{n+1}a_j\Big)-\exp\Big(\sum_{j=1}^{N_1}a_j\Big)\Big)$$
$$\leqslant o(1)+4\varepsilon+3\varepsilon e^{a_{n+1}}. \tag{3.5.65}$$

所以 $\limsup\limits_{n\to\infty}\|u_n\|\leqslant 7\varepsilon$，由此即得式(3.5.60)。

当 $n\geqslant N_2$ 时，$\|x_n-u_n\|\leqslant M_\sigma+8$，所以

$$h^\tau(x_n)v_x(x_n)I_{[\|x_n-x^0\|\geqslant\rho]}\leqslant-\alpha(\rho)L_0I_{[\|x_n-x^0\|\geqslant\rho]}$$
$$\leqslant-\alpha(\rho)v(x_n-u_n)I_{[\|x_n-x^0\|\geqslant\rho]}.$$

对 $n\geqslant N_2$，由式(3.5.54)～(3.5.57)及 Taylor 展式得

$$v(x_{n+1}-u_{n+1})=v(x_n-u_n)+a_n(h^\tau(x_n)+u_n^\tau)v_x(x_n-u_n)$$
$$+\frac{1}{2}a_n^2(h^\tau(x_n)+u_n^\tau)v_{xx}(\zeta_n)(h(x_n)+u_n)$$
$$\leqslant v(x_n-u_n)+a_nh^\tau(x_n)(v_x(x_n)-v_{xx}(\eta_n)u_n)+a_n8L_1\varepsilon$$
$$+L_2(L_3^2+64\varepsilon^2)a_n^2$$
$$\leqslant(1-\alpha(\rho)a_n)v(x_n-u_n)$$
$$+a_n(0\vee h^\tau(x_n)v_x(x_n))I_{[\|x_n-x^0\|<\rho]}$$
$$+\alpha(\rho)a_nv(x_n-u_n)I_{[\|x_n-x^0\|<\rho]}+8\varepsilon a_nL_2L_3$$
$$+a_n8L_1\varepsilon+L_2(L_3^2+64\varepsilon^2)a_n^2. \tag{3.5.66}$$

当 $\|x_n-x^0\|<\rho$ 时，估 $v(x_n-u_n)$。

$$v(x_n-u_n)I_{[\|x_n-x^0\|<\rho]}$$
$$=[v(x^0)+v_x(\eta)(x_n-u_n-x^0)]\cdot I_{[\|x_n-x^0\|<\rho]},$$

因为 $\|x_n-x^0\|<\rho$，所以

$$\|x^0\|\leqslant\|x^0-x_n+x_n\|\leqslant M_\sigma+\rho,$$

因此 $\|\eta\|\leqslant M_\sigma+8$，$\|v_x(\eta)\|\leqslant L_1$，所以有

$$v(x_n-u_n)I_{[\|x_n-x^0\|<\rho]}\leqslant L_1(\rho+8\varepsilon), \tag{3.5.67}$$

记 $L_4=8(L_2L_3+L_1)$，$L_5=L_2(L_3^2+64\varepsilon^2)$，

从式(3.5.66)、(3.5.67)得

$$
\begin{aligned}
v(x_{n+1}-u_{n+1}) &\leqslant (1-\alpha(\rho)a_n)v(x_n-u_n) \\
&\quad + a_n(0\vee\beta(\rho)) + L_1(\rho+8\varepsilon)\alpha(\rho)a_n \\
&\quad + L_4\varepsilon a_n + L_5 a_n^2 \leqslant \exp v(x_{N_2}-u_{N_2})\left[-\sum_{i=N_2}^{n}\alpha(\rho)a_i\right] \\
&\quad + \sum_{i=N_2}^{n}\prod_{j=i+1}^{n}(1-\alpha(\rho)a_j)a_i((0\vee\beta(\rho)) \\
&\quad + L_1(\rho+8\varepsilon)\alpha(\rho) + L_4\varepsilon + L_5 a_i).
\end{aligned}
$$

和式(3.5.65)一样类似地证明

$$
\begin{aligned}
\sum_{i=N_2}^{n}\prod_{j=i+1}^{n}(1-\alpha(\rho)a_j)a_i &\leqslant \frac{3}{2}\sum_{i=N_2}^{n}a_{i+1}\exp\left[-\sum_{j=i+1}^{n}\alpha(\rho)a_j\right] \\
&\leqslant \frac{3}{2\alpha(\rho)}\exp\left[-\sum_{j=1}^{n}\alpha(\rho)a_j\right]\sum_{i=N_2}^{n}(e^{\alpha(\rho)a_{i+1}}-1)\exp\left[\sum_{j=1}^{i}\alpha(\rho)a_j\right] \\
&\leqslant \frac{3}{2\alpha(\rho)}e^{\alpha(\rho)a_{n+1}}.
\end{aligned}
$$

所以

$$
\begin{aligned}
v(x_{n+1}-u_{n+1}) &\leqslant v(x_{N_2}-u_{N_2})\exp\left[-\sum_{i=N_2}^{n}\alpha(\rho)a_i\right] \\
&\quad + \frac{3}{2\alpha(\rho)}e^{\alpha(\rho)a_{n+1}}((0\vee\beta(\rho)) + L_4\varepsilon + L_5 a_{N_2}) \\
&\quad + \frac{3}{2}e^{\alpha(\rho)a_{n+1}}L_1(\rho+8\varepsilon) \triangleq t_n.
\end{aligned}
\tag{3.5.68}
$$

定义

$$
m(r) = \min_{\|x-x^0\|\geqslant r} v(x), \quad r\geqslant 0, \tag{3.5.69}
$$

$m(r)$是非降函数, $m(r)\xrightarrow[r\to 0]{} m(0)=0$.

取 r_n, 使 $m(r_n)=2t_n$, 那么

$$
v(x_{n+1}-u_{n+1}) \leqslant \frac{m(r_n)}{2} < m(r_n). \tag{3.5.70}
$$

定义函数 $f(\cdot)$:

$$
f(t) = \min\{r: m(r)=t\},
$$

显然 $f(\cdot)$左连续, 非降, $f(t)\xrightarrow[t\to 0^+]{} 0$.

从式(3.5.69)及式(3.5.70)可知

$$\|x_{n+1}-u_{n+1}-x^0\|<r_n=f(2t_n).$$

用式(3.5.60)及 t_n 的定义，从上式知

$$\limsup_{n\to\infty}\|x_n-x^0\|$$

$$<f\left(\frac{3}{\alpha(\rho)}((0\vee\beta(\rho))+L_4\varepsilon)+3L_1(\rho+8\varepsilon)\right)+8\varepsilon. \quad \square$$

第 4 章

随机逼近算法的渐近性质

在第 3 章，我们已比较充分地讨论了随机逼近算法的几乎处处收敛性，本章将来讨论算法的渐近正态性及渐近有效性. 估计的渐近分布对假设检验很重要，而使估计误差的协方差阵渐近地达到最小的估计，自然受到特别重视. 一维随机逼近算法的渐近正态性见文献[26]，多维的见文献[38, 83]. 以后不少学者继续从事这一问题的研究，将以上结果推广到相关量测误差[31, 49, 84]、连续时间情形[20]和非线性量测误差等[51, 52, 87, 88, 89].

钟开莱在文献[26]中还指出，如果我们能知道回归函数在零点的导数(多维情况时即梯度阵)$A=h_x(x^0)$，可以取步长因子$a_n=\frac{1}{n}A^{-1}$，这样得到的 RM 算法的渐近方差是最小的，也就是说算法是有渐近有效性. 但是联系到随机逼近算法的问题本身，A 阵不能预先知道，因而这样的算法实际上不好实现. 文献[86]构造了一种适应性的随机逼近算法，在每一步利用观测值构造出 A 的差商估计 A_n，然后令 $a_n=\frac{1}{n}A_n^{-1}$，最后证明了新算法的渐近有效性. 到 1987 年以后，这个结果又被推广到多维情况，参见文献[91].

这种适应性随机逼近算法虽然能实现算法的渐近有效性，但所要求的条件过强，并且构造差商要增加观测量，特别在高维情况计算量大增. 此外还要考虑 A_n 可能退化而引起算法发散的问题. 由于这些缺点，人们希望有其他办法来实现渐近有效性.

直到 80 年代末，文献[76, 77, 82]采用衰减速度慢于$\frac{1}{n}$的步

长因子，例如 $\frac{1}{n^\alpha}$，$\alpha<1$，使算法的收敛速度减慢(见§4.2)，然后再对估计取平均，实现了渐近有效性．进一步的研究见文献[24, 25, 60, 61, 95, 96]．

§4.1 线性递推估计的渐近正态性

在上一章§3.2中我们看到，当回归函数 $h(\cdot)$ 满足条件(3.2.1)时，变界截尾算法(3.1.3)、(3.1.4)的渐近行为取决于线性递推方程(3.2.9)的渐近性质．这一节我们来讨论这类递推估计的渐近正态性，也是为下一节做准备．应该注意，在讨论渐近正态性时，步长因子不一定限于 $\frac{a}{n}$ 的形式．

我们要用一个双指标随机矢量的中心极限定理，记为引理4.1.1，它的证明可参见文献[74]．

引理 4.1.1 设 ξ_{nk}，$1\leqslant k\leqslant n$，是 l 维随机矢量列．

记 $S_{nk}=E\xi_{nk}\xi_{nk}^\tau$，$R_{nk}=E(\xi_{nk}\xi_{nk}^\tau|\xi_{n1},\cdots,\xi_{nk-1})$，

$$S_n=\sum_{k=1}^n S_{nk}.$$

假设

$$E(\xi_{nk}|\xi_{n1},\cdots,\xi_{nk-1})=0, \tag{4.1.1}$$

$$\sup_{n\geqslant 1}\sum_{k=1}^n E\|\xi_{nk}\|^2<\infty,\quad \lim_{n\to\infty}S_n=S, \tag{4.1.2}$$

$$\lim_{n\to\infty}\sum_{k=1}^n E\|S_{nk}-R_{nk}\|=0, \tag{4.1.3}$$

$$\lim_{n\to\infty}\sum_{k=1}^n E\|\xi_{nk}\|^2 I_{[\|\xi_{nk}\|>\varepsilon]}=0,\quad \forall\varepsilon>0, \tag{4.1.4}$$

则

$$\sum_{k=1}^n \xi_{nk}\xrightarrow[n\to\infty]{d}N(0,S). \tag{4.1.5}$$

以后我们总用 $N(0,S)$ 表示期望为零、协方差阵为 S 的正态分布．

现在来看 l 维线性递推估计(3.2.9)．

$$u_{n+1}=(I+a_nH_n)u_n+a_n(e_{n+1}+\nu_{n+1}),\ 初值\ u_0. \quad (4.1.6)$$

对步长因子，H_n 及噪声，我们要用以下条件：

A 4.1.1　　$a_n>0,\ a_n\to 0,\ \sum_{i=1}^{\infty}a_i=\infty,\ a_{n+1}^{-1}-a_n^{-1}\xrightarrow[n\to\infty]{}\alpha\geqslant 0.$

(4.1.7)

A 4.1.2　$H_n\xrightarrow[n\to\infty]{}H,\ H+\frac{1}{2}\alpha I$ 为稳定阵.

A4.1.3　$\nu_n=o(\sqrt{a_n}),\ e_n=\sum_{i=0}^{\infty}C_i\omega_{n-i},\quad \omega_i=0,\ i<0,$ (4.1.8)

C_i 是 $l\times l$ 常阵，$\sum_{i=0}^{\infty}\|C_i\|<\infty$，而 $\{\omega_i,\ \mathscr{F}_i\}$ 是 l 维鞅差序列，满足

$$E(\omega_i|\mathscr{F}_{i-1})=0,\ E(\|\omega_i\|^2|\mathscr{F}_{i-1})\leqslant\sigma,\ \forall i(\sigma\ 为常数),$$

(4.1.9)

$$\lim_{n\to\infty}E(\omega_i\omega_i^{\tau}|\mathscr{F}_{i-1})=S_0,\ \text{a. s.}(S_0\ 为常阵), \quad (4.1.10)$$

$$\lim_{N\to\infty}\sup_i E\|\omega_i\|^2I_{[\|\omega_i\|>N]}=0. \quad (4.1.11)$$

容易验证，当 $a_n=\frac{a}{n}$ 时，$a_{n+1}^{-1}-a_n^{-1}=\frac{1}{a}$；当 $a_n=\frac{a}{n^{\beta}}$ 时，$\frac{1}{2}<\beta<1$，$\alpha=0$，它们都使 A 4.1.1 成立．此外，$a_n=\frac{\log n}{n^{\beta}}$，$\beta\leqslant 1$；$a_n=\frac{1}{n^{\beta}\log n}$，$\beta<1$ 也使 A 4.1.1 成立．

如果对某个 $\theta>2$

$$E(\|\omega_i\|^{\theta}|\mathscr{F}_{i-1})\leqslant k,\ \forall i,$$

k 为一个常数，那么用 Hölder 不等式$\left(\frac{2}{\theta}+\frac{\theta-2}{\theta}=1\right)$和 Markov 不等式，就有

$$\begin{aligned}
&\lim_{N\to\infty}\sup_n \mathrm{E}\|\omega_n\|^2I_{[\|\omega_n\|>N]}\\
&\leqslant\lim_{N\to\infty}\sup_n \mathrm{E}(\|\omega_n\|^{\theta})^{\frac{2}{\theta}}(\mathrm{E}(I_{[\|\omega_n\|>N]})^{\frac{\theta}{\theta-2}})^{\frac{\theta-2}{\theta}}\\
&\leqslant\lim_{N\to\infty}\sup_n k^{\frac{2}{\theta}}(\mathrm{E}(I_{[\|\omega_n\|>N]}))^{\frac{\theta-2}{2}}\\
&\leqslant\lim_{N\to\infty}\sup_n k^{\frac{2}{\theta}}\ (P(\|\omega_n\|\geqslant N))^{\frac{\theta-2}{\theta}}
\end{aligned}$$

$$\leqslant \lim_{N\to\infty} \sup_n k^{\frac{2}{\delta}} \left(\frac{\mathrm{E}\|\omega_n\|^2}{N^2} \right)^{\frac{\delta-2}{\delta}}$$

$$\leqslant \lim_{N\to\infty} k^{\frac{2}{\delta}} \cdot \frac{k^{\frac{2(\delta-2)}{\delta^2}}}{N^{\frac{2(\delta-2)}{\delta}}} = 0.$$

所以这时，式(4.1.11)成立.

仍沿用§3.2中引进的符号，记

$$\begin{cases} \Phi_{n,k} = (I + a_n H_n) \cdots (I + a_k H_k) \triangleq \prod_{m=k}^{n} (I + a_m H_m), \quad k \leqslant n, \\ \Phi_{j,j+1} = I, \end{cases} \tag{4.1.12}$$

当 $H_n \xrightarrow[n\to\infty]{} H$，$H$ 为稳定阵时，从式(3.2.11)知

$$\|\Phi_{n,k}\| \leqslant c_0 \exp\left(-c \sum_{i=k}^{n} a_i\right),\ c_0,\ c>0 \text{ 为常数}. \tag{4.1.13}$$

进一步可有以下结果.

引理4.1.2　设 $a_n>0$，$a_n\to 0$，$\sum_{n=1}^{\infty} a_n=\infty$，$H_n \xrightarrow[n\to\infty]{} H$，$H$ 为稳定阵，那么对任意 $r>0$，

$$\sup_n \sum_{k=1}^{n} a_k \|\Phi_{n,k+1}\|^r < \infty. \tag{4.1.14}$$

证明　由式(3.2.11)知

$$\sum_{k=1}^{n} a_k \|\Phi_{n,k+1}\|^r \leqslant c_0^r \sum_{k=1}^{n} a_k \exp\left(-cr \sum_{i=k+1}^{n} a_i\right), \tag{4.1.15}$$

注意到 $1-e^{-x}-x+\frac{x^2}{2}$，当 $x=0$ 时为0，而它的导数 $e^{-x}-1+x\geqslant 0$，所以 $1-e^{-x}+\frac{x^2}{2}\geqslant x$，$\forall x\geqslant 0$，因此，

$$\sum_{k=1}^{n} a_k \exp\left(-cr \sum_{i=k+1}^{n} a_i\right)$$

$$\leqslant \frac{1}{cr} \sum_{k=1}^{n} \left(1-e^{-cra_k}+\frac{(cra_k)^2}{2}\right) \exp\left(-cr \sum_{i=k+1}^{n} a_i\right)$$

$$= \frac{1}{cr} \sum_{k=1}^{n} \left[\exp\left(-cr \sum_{i=k+1}^{n} a_i\right) - \exp\left(-cr \sum_{i=k}^{n} a_i\right)\right]$$

$$+\frac{cr}{2}\sum_{k=1}^{n}a_k^2\exp\left(-cr\sum_{i=k+1}^{n}a_i\right),$$

或者
$$\sum_{k=1}^{n}\left(1-\frac{cr}{2}a_k\right)a_k\exp\left(-cr\sum_{i=k+1}^{n}a_i\right)\leqslant\frac{1}{cr},$$

对充分大的k_1，有$\frac{cr}{2}a_k<\frac{1}{2}$，$\forall k\geqslant k_1$，所以

$$\begin{aligned}&\sum_{k=1}^{n}a_k\exp\left(-cr\sum_{i=k+1}^{n}a_i\right)\\&\leqslant\sum_{k=1}^{k_1-1}a_k+2\sum_{k=k_1}^{n}\left(1-\frac{cr}{2}a_k\right)a_k\exp\left(-cr\sum_{i=k+1}^{n}a_i\right)\\&\leqslant\sum_{k=1}^{k_1-1}a_k+\frac{2}{cr},\end{aligned}\tag{4.1.16}$$

与式(4.1.15)一起就证明了式(4.1.14).

引理4.1.3 在引理4.1.2的条件下，记

$$\theta_{n,k}=\exp\left(H\sum_{i=k}^{n}a_i\right)-\Phi_{n,k},$$

那么$\sup\limits_{n,k}\|\theta_{n,k}\|<\infty$，并且$\|\theta_{n,k}\|\xrightarrow[k\to\infty]{}0$，对$n\geqslant k$一致，对固定的$k$，$\|\theta_{n,k}\|\xrightarrow[k\to\infty]{}0$.

证明 用Taylor展式知，$e^{a_nH}=(I+a_nH+A_n)$，$\|A_n\|=O(a_n^2)$，所以

$$\begin{aligned}\theta_{n,k}&=(I+a_nH+A_n)\exp\left(H\sum_{i=k}^{n-1}a_i\right)-(I+a_nH_n)\Phi_{n-1,k}\\&=(I+a_nH_n)\theta_{n-1,k}+(a_n(H-H_n)+A_n)\exp\left(H\sum_{i=k}^{n-1}a_i\right)\\&=\sum_{j=k}^{n}\Phi_{n,j+1}(a_j(H-H_j)+A_j)\exp\left(H\sum_{i=k}^{j-1}a_i\right),\end{aligned}\tag{4.1.17}$$

这里
$$\sum_{i=k}^{k-1}a_i\triangleq0.$$

因H稳定，所以存在常数$c_1>0$，$\rho>0$，使

$$\left\|\exp\left(H\sum_{i=k}^{n}a_i\right)\right\|\leqslant c_1\exp\left(-\rho\sum_{i=k}^{n}a_i\right),\tag{4.1.18}$$

从式(4.1.17)、(4.1.18)知

$$\|\theta_{n,k}\| \leqslant \sum_{j=k}^{n} \|\Phi_{n,j+1}\| a_j (o(1)+O(a_j)) c_1 \exp\left(-\rho \sum_{i=k}^{j-1} a_i\right)$$

$$\leqslant c_1 \sum_{j=k}^{n} a_j \|\phi_{n,j+1}\| (o(1)+O(a_j)),$$

对任意 $\varepsilon>0$, 当 $j\geqslant k_1$ 时, $(o(1)+O(a_j))<\frac{\varepsilon}{2}$, 所以由式(4.1.13)知

$$\|\theta_{n,k}\| \leqslant \frac{\varepsilon}{2} c_1 \sum_{j=k_1}^{n} a_j \|\Phi_{n,j+1}\|$$

$$+c_0 c_1 \sum_{j=k}^{k_1-1} a_j \left[\exp\left(-c \sum_{i=j+1}^{n} a_i\right)\right] (o(1)+O(a_j)).$$

注意到 $\sum_{i=1}^{\infty} a_i=\infty$, 利用引理 4.1.2,由此立即得出引理结论.□

引理 4.1.4 设 $a_n>0$, $a_n\to 0$, $\sum_{n=1}^{\infty} a_n=\infty$, 设 A, B, Q 都是 $l\times l$ 阵, A 和 B 稳定, 那么

$$\lim_{n\to\infty} \sum_{i=1}^{n} a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q \exp\left(B \sum_{j=i+1}^{n} a_j\right)$$

$$=\int_0^{\infty} e^{At} Q e^{Bt}\, \mathrm{d}t. \tag{4.1.19}$$

证明 对任意 $T>0$, 定义

$$s(n,T)=\min\left\{s: \sum_{i=s}^{n} a_i \leqslant T\right\}, \tag{4.1.20}$$

由于 $a_i\to 0$, 所以对固定的 T, $\sum_{i=s(n,T)}^{n} a_i \to T$. 所以如果把 $\sum_{j=s(n,T)}^{i} a_j$ 记作 t, 那么 $\sum_{j=i+1}^{n} a_j \xrightarrow[n\to\infty]{} T-t$, 所以 $\sum_{i=s(n,T)}^{n} a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q$ $\cdot\exp\left(B \sum_{j=i+1}^{n} a_j\right)$ 是 $\int_0^T e^{A(T-t)} Q e^{B(T-t)}\, \mathrm{d}t$ 的积分和,也是 $\int_0^T e^{At} Q e^{Bt} \mathrm{d}t$ 的积分和, 因此

$$\lim_{n\to\infty} \sum_{i=s(n,T)}^{n} a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right) Q\right] \exp\left(B \sum_{j=i+1}^{n} a_j\right)$$

$$=\int_0^T e^{At} Q e^{Bt}\, \mathrm{d}t, \tag{4.1.21}$$

注意到 A 稳定，和式 (4.1.16) 类似地知存在常数 $\eta_0<\infty$，$\sup_n \sum_{i=1}^{n} a_i \left\| \exp\left(A \sum_{j=i+1}^{n} a_j\right) \right\| <\eta_0$，所以

$$\sup_{n,T} \sum_{i=1}^{s(n,T)-1} a_i \left\| \exp\left(A \sum_{j=i+1}^{s(n,T)-1} a_j\right) \right\| <\eta_0.$$

由于 B 稳定，所以存在常数 η_1，使 $\left\| Q \exp\left(B \sum_{j=i+1}^{n} a_i\right) \right\| \leqslant \eta_1$，$\forall n\geqslant i+1$. 因为 A 稳定，所以 $\|e^{At}\|\leqslant \eta_2 e^{-\eta t}$，$\forall t\geqslant 0$，$\eta_2>0$，$\eta>0$. 因此

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{i=1}^{s(n,T)-1} a_i \left\| \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q \exp\left(B \sum_{j=i+1}^{n} a_j\right) \right\| \\
&\leqslant \eta_1 \lim_{n\to\infty} \left\| \exp\left(A \sum_{j=s(n,T)}^{n} a_j\right) \right\| \sum_{i=1}^{s(n,T)-1} a_i \left\| \exp\left(A \sum_{j=i+1}^{s(n,T)-1} a_j\right) \right\| \\
&\leqslant \eta_0\eta_1\eta_2 \lim_{n\to\infty} e^{-\eta(T-a_{s(n,T)-1})} = \eta_0\eta_1\eta_2 e^{-\eta T},
\end{aligned} \tag{4.1.22}$$

最后一个等式是由于对固定的 T，$s(n, T)\xrightarrow[n\to\infty]{}\infty$.

由式(4.1.21)、(4.1.22)便知

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{i=1}^{n} a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q \exp\left(B \sum_{j=i+1}^{n} a_j\right) \\
&= \lim_{T\to\infty} \lim_{n\to\infty} \sum_{i=1}^{n} a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q \exp\left(B \sum_{j=i+1}^{n} a_j\right) \\
&= \lim_{T\to\infty} \lim_{n\to\infty} \left(\sum_{i=1}^{s(n,T)-1} + \sum_{i=s(n,T)}^{n}\right) a_i \left[\exp\left(A \sum_{j=i+1}^{n} a_j\right)\right] Q \exp\left(B \sum_{j=i+1}^{n} a_j\right) \\
&= \int_0^\infty e^{At} Q e^{Bt}\, dt.
\end{aligned}$$

□

现证由式(4.1.6)给出的 u_n 的渐近正态性[105].

定理4.1.1 设条件 A 4.1.1～A 4.1.3 成立，那么 $\frac{1}{\sqrt{a_n}} u_n$ 渐近正态，

$$\frac{1}{\sqrt{a_n}} u_n \to N(0, S),$$

$$S = \int_0^\infty e^{\left(H+\frac{1}{2}aI\right)t} \sum_{i=0}^{\infty} C_i s_0 \sum_{i=0}^{\infty} C_i^\tau e^{\left(H^\tau+\frac{1}{2}aI\right)t}\, dt.$$

证明 设 u_n' 由下列递推式定义：

$$u'_{n+1}=(I+a_nH_n)u'_n+a_ne_{n+1},\ u'_1=0,\tag{4.1.23}$$

那么 $$u_{n+1}-u'_{n+1}=(I+a_nH_n)(u_n-u'_n)+a_nv_{n+1},$$

由式(4.1.7)知

$$\begin{aligned}\left(\frac{a_n}{a_{n+1}}\right)^{\frac{1}{2}}&=\left(\frac{a_n-a_{n+1}}{a_{n+1}}+1\right)^{\frac{1}{2}}\\&=1+\frac{1}{2}a_n(a_{n+1}^{-1}-a_n^{-1})+O\left(\left(\frac{a_n-a_{n+1}}{a_{n+1}}\right)^2\right)\\&=1+\frac{1}{2}a_n\alpha+o(a_n).\end{aligned}\tag{4.1.24}$$

所以

$$\begin{aligned}&\frac{u_{n+1}-u'_{n+1}}{\sqrt{a_{n+1}}}\\&=\left(1+\frac{1}{2}a_n\alpha+o(a_n)\right)\left[(I+a_nH_n)\frac{(u_n-u'_n)}{\sqrt{a_n}}+\sqrt{a_n}v_{n+1}\right]\\&=(I+a_nA_n)\frac{u_n-u'_n}{\sqrt{a_n}}+\sqrt{a_n}v'_{n+1},\end{aligned}$$

这里

$$\begin{aligned}A_n&=H_n+\frac{1}{2}\alpha I+\frac{1}{2}a_n\alpha H_n+o(\sqrt{a_n})(I+a_nH_n)\\&\to A\triangleq H+\frac{1}{2}\alpha I,\end{aligned}$$

而用 A 4.1.3 知

$$v'_{n+1}\triangleq\left(1+\frac{1}{2}a_n\alpha+o(a_n)\right)v_{n+1}=o(\sqrt{a_{n+1}}).\tag{4.1.25}$$

据 A 4.1.2, A 稳定, 所以记

$$\psi_{n,k}=(I+a_nA_n)\cdots(I+a_kA_k),\ \psi_{k-1,k}=I,\tag{4.1.26}$$

从式(3.2.11)知, 存在常数 $b_0>0$, $b>0$, 使

$$\|\psi_{n,k}\|\leqslant b_0\exp\left(-b\sum_{i=k}^{n}a_i\right),\ \forall n\geqslant k,\tag{4.1.27}$$

那么

$$\frac{u_{n+1}-u'_{n+1}}{\sqrt{a_{n+1}}}=\psi_{n,1}\left(\frac{u_1-u'_1}{\sqrt{a_1}}\right)+\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\,v'_{i+1},\tag{4.1.28}$$

由式(4.1.27)知，式(4.1.28)中左端第一项当 $n\to\infty$ 时趋于 0，用引理 4.1.2 及式(4.1.27)，并注意到式(4.1.25)便知式(4.1.28)最后一项也趋于 0，所以 $\frac{u_n}{\sqrt{a_n}}$ 和 $\frac{u'_n}{\sqrt{a_n}}$ 的极限分布相同. 所以只要证明

$$\frac{u'_n}{\sqrt{a_n}}\xrightarrow[n\to\infty]{d}N(0,\ S).$$

由式(4.1.24)和式(4.1.28)类似地得到

$$\begin{aligned}\frac{1}{\sqrt{a_{n+1}}}u'_{n+1}=&\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=0}^{\infty}C_j\omega_{i+1-j}\\&+\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}e_{i+1}\Big(\frac{1}{2}a_i\alpha+o(a_i)\Big),\end{aligned}\tag{4.1.29}$$

由式(4.1.8)及式(4.1.11)易见 $E\|e_i\|^2$ 对 i 一致有界、由于 A_n 趋于一个稳定阵，由式(3.2.11)及引理 4.1.2，很容易看出

$$E\left\|\sum_{i=1}^{n}\psi_{n,i+1}a_ie_{i+1}\Big(\frac{1}{2}\alpha\sqrt{a_i}+o(\sqrt{a_i})\Big)\right\|\xrightarrow[n\to\infty]{}0,\tag{4.1.30}$$

所以式(4.1.29)中最后一项依概率收敛到 0，故 $\frac{u_{n+1}}{\sqrt{a_{n+1}}}$ 的极限分布应和 $\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=0}^{\infty}C_j\omega_{i+1-j}$ 的极限分布相同. 换句话说，我们只要证明

$$\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=0}^{\infty}C_j\omega_{i+1-j}\xrightarrow[n\to\infty]{d}N(0,\ s).\tag{4.1.31}$$

固定 $r>0$，那么

$$\begin{aligned}&\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=0}^{\infty}C_j(\omega_{i+1-j}-\omega_{i+1})\\&=\sum_{j=0}^{r}\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}C_j(\omega_{i+1-j}-\omega_{i+1})\\&\quad+\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=r+1}^{\infty}C_j(\omega_{i+1-j}-\omega_{i+1}).\end{aligned}\tag{4.1.32}$$

由式(4.1.9)知

$$E\left\|\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\sum_{j=r+1}^{\infty}C_j\omega_{i+1-j}\right\|$$

$$\leqslant \sum_{j=r+1}^{\infty}\|C_j\| E\left\|\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\,\omega_{i+1-j}\right\|$$

$$\leqslant \sum_{j=r+1}^{\infty}\|C_j\|\left(\sum_{i=1}^{n}\|\psi_{n,i+1}\|^2 a_i\sigma\right)^{\frac{1}{2}}.$$

由引理 4.1.2 及 $\sum\limits_{j=0}^{\infty}\|C_j\|<\infty$, 便知上式右端当 $r\to\infty$ 时，对 n 一致地趋于 0，所以式(4.1.32)式中最后一项当 $r\to\infty$ 时依概率趋于 0, 而式(4.1.32)右端第一项可表达为

$$\sum_{j=0}^{r}\sum_{i=1}^{j}\sqrt{a_i}\,\psi_{n,i+1}C_j\omega_{i+1-j}-\sum_{j=0}^{r}\sum_{i=n-j+1}^{n}\sqrt{a_i}\,\psi_{n,i+1}C_j\omega_{i+1}$$

$$-\sum_{j=0}^{r}\sum_{i=1}^{n-j}(\sqrt{a_i}\,\psi_{n,i+1}-\sqrt{a_{i+j}}\,\psi_{n,i+j+1})C_j\omega_{i+1} \tag{4.1.33}$$

注意到 r 是固定的，对 $\psi_{n,i+1}$ 由式(3.2.11)型的估计，由于 $\sum\limits_{i=1}^{\infty}a_i=\infty$, 所以式(4.1.33)的第一项 a. s. 趋于 0. 又由于 r 固定，而 $a_n\xrightarrow[n\to\infty]{}0$, 所以对式(4.1.33)的第二项的范数求期望后，知它依概率收敛到 0.

现在来看式(4.1.33)的最后一项.

首先，从式(4.1.24)知，对固定的 $j\geqslant1$, $\frac{a_n}{a_{n+j}}\xrightarrow[n\to\infty]{}1$, 并且 $\left(\frac{a_n}{a_{n+j}}\right)^{\frac{1}{2}}=1+O(a_n)$, 所以，对固定的 r, 存在常数 m_0,

$$0\leqslant\left(\frac{a_n}{a_{n+j}}\right)^{\frac{1}{2}}-1\leqslant m_0a_{n+j},\quad \forall n\geqslant1,\ \forall j\geqslant r. \tag{4.1.34}$$

所以

$$\sum_{j=0}^{r}\sum_{i=1}^{n-j}(\sqrt{a_i}\,\psi_{n,i+1}-\sqrt{a_{i+j}}\,\psi_{n,i+j+1})C_j\omega_{i+1}$$

$$=\sum_{j=0}^{r}\sum_{i=1}^{n-j}\sqrt{a_i}\,(\psi_{n,i+1}-\psi_{n,i+j+1})C_j\omega_{i+1}$$

$$+\sum_{j=0}^{r}\sum_{i=1}^{n-j}\left(\left(\frac{a_i}{a_{i+j}}\right)^{\frac{1}{2}}-1\right)\sqrt{a_{i+j}}\,\psi_{n,i+j}C_j\omega_{i+1}$$

$$=\sum_{j=0}^{r}\sum_{i=1}^{n-j}\sqrt{a_i}\sum_{k=1}^{j}\psi_{n,i+k+1}a_{i+k}A_{i+k}C_j\omega_{i+1}$$

$$+m_0\sum_{j=0}^{r}\sum_{i=1}^{n-j}\left(\left(\frac{a_i}{a_{i+j}}\right)^{\frac{1}{2}}-1\right)\sqrt{a_{i+j}}\,\psi_{n,i+j}C_j\omega_{i+1}. \quad (4.1.35)$$

对固定的 r, 存在常数 m_1, 使 $\sum_{k=1}^{j}\|A_{i+k}C_j\|^2\sigma\leqslant m_1$ $\forall j\leqslant r$, $\forall k\leqslant r$, $\forall i\geqslant 1$, 所以

$$\sum_{j=0}^{r}E\left\|\sum_{i=1}^{n-j}\sqrt{a_i}\sum_{k=1}^{j}\psi_{n,i+k+1}a_{i+k}A_{i+k}C_j\omega_{i+1}\right\|$$

$$\leqslant\sum_{j=0}^{r}\left(\sum_{i=1}^{n-j}a_i\left\|\sum_{k=1}^{j}\psi_{n,i+k+1}a_{i+k}A_{i+k}C_j\right\|^2\sigma\right)^{\frac{1}{2}}$$

$$\leqslant\sum_{j=0}^{r}\left(\sum_{i=1}^{n-j}a_i\sum_{k=1}^{j}\|\psi_{n,i+k+1}\|^2a_{i+k}^2\sum_{k=1}^{j}\|A_{i+k}C_j\|^2\sigma\right)^{\frac{1}{2}}$$

$$\leqslant\sum_{j=0}^{r}\left(m_1\sum_{i=1}^{n-j}a_i\sum_{k=1}^{j}\|\psi_{n,i+k+1}\|^2a_{i+k}^2\right)^{\frac{1}{2}}$$

$$\leqslant\sum_{j=0}^{r}\left(m_1\sum_{k=1}^{j}\sum_{i=1}^{n-j}a_i\|\psi_{n,i+k+1}\|^2a_{i+k}^2\right)^{\frac{1}{2}}\xrightarrow[n\to\infty]{}0, \quad (4.1.36)$$

上式趋于 0, 因为对任一$j\leqslant r$, $\sum_{i=1}^{n_1}a_i\|\psi_{n,i+k+1}\|^2a_{i+k}^2\xrightarrow[n\to\infty]{}0$, 而当 $i\geqslant n_1$ 时, a_i 可任意小, 所以用引理 4.1.2 知式(4.1.36)正确. 所以对任意固定的 r, 式(4.1.35)右端第一式依概率趋于 0. 类似地用引理 4.1.2 及式(4.1.34)知, 式(4.1.35)的最后一式也依概率趋于 0. 所以式 (4.1.33) 三项全部依概率收敛到 0. 这样, 我们在式(4.1.32)中首先把 $n\to\infty$, 然后把 $r\to\infty$, 便知式(4.1.32)右端依概率收敛到 0, 这样从式(4.1.31)知, 我们只要证明

$$\sum_{i=1}^{n}\psi_{n,i+1}\sqrt{a_i}\left(\sum_{j=0}^{\infty}C_j\right)\omega_{i+1}\xrightarrow[n\to\infty]{d}N(0,\ s). \quad (4.1.37)$$

我们现在把 $\psi_{n,i+1}\sqrt{a_i}\left(\sum_{j=0}^{\infty}C_j\right)\omega_{i+1}$ 等同引理 4.1.1 中的 $\xi_{n,i}$, 并验证引理的条件成立.

因为$\{\omega_i, \mathscr{F}_i\}$是鞅差列, 所以式 (4.1.1) 显然成立. 条件(4.1.10), 引理 4.1.2 及式(4.1.27)保证了式(4.1.3)成立. 引理 4.1.2 保证式(4.1.2)中第一个要求成立, 现在来验证 $\lim\limits_{n\to\infty}S_n=S$.

$$S_n=\sum_{i=1}^{n}a_i\psi_{n,i+1}\left(\sum_{j=0}^{\infty}C_j\right)E\omega_{i+1}\omega_{i+1}^{\tau}\left(\sum_{j=0}^{\infty}C_j\right)^{\tau}\psi_{n,i+1}^{\tau}. \tag{4.1.38}$$

由式(4.1.10)、引理4.1.2及式(4.1.27)知

$$\left\|S_n-\sum_{i=1}^{n}a_i\psi_{n,i+1}\left(\sum_{j=0}^{\infty}C_j\right)S_0\left(\sum_{j=0}^{\infty}C_j\right)^{\tau}\psi_{n,i+1}^{\tau}\right\|\xrightarrow[n\to\infty]{}0 \tag{4.1.39}$$

在引理4.1.3中把$\Phi_{n,k}$换成$\psi_{n,k}$，相应地H换成$A=H+\frac{1}{2}\alpha I$. 引理结论照样成立. 那么用该引理及式(4.1.27)知

$$\sum_{i=1}^{n}a_i\left\|\exp\left(A\sum_{j=i+1}^{n}a_j\right)-\psi_{n,i+1}\right\|\|S_0\|\left\|\sum_{j=0}^{\infty}C_j\right\|^2\|\psi_{n,i+1}\|\xrightarrow[n\to\infty]{}0, \tag{4.1.40}$$

所以我们就有

$$\left\|S_n-\sum_{i=1}^{n}a_i\left[\exp\left(A\sum_{j=i+1}^{n}a_j\right)\right]\left(\sum_{j=0}^{\infty}C_j\right)S_0\left(\sum_{j=0}^{\infty}C_j\right)^{\tau}\cdot\exp\left(A^{\tau}\sum_{j=i+1}^{n}a\right)\right\|\xrightarrow[n\to\infty]{}0, \tag{4.1.41}$$

用引理4.1.4便知

$$\|S_n-S\|\xrightarrow[n\to\infty]{}0.$$

最后，我们验证式(4.1.4).

由式(4.1.27)知

$$\left[\left\|\psi_{n,i+1}\sqrt{a_i}\left(\sum_{j=0}^{\infty}C_j\right)\omega_{i+1}\right\|>\varepsilon\right]\subset\left[\|\omega_{i+1}\|>(b_0)^{-1}\left\|\sum_{j=0}^{\infty}C_j\right\|^{-1}\frac{\varepsilon}{\sqrt{a_i}}\exp\left(b\sum_{j=i+1}^{n}a_j\right)\right],$$

由于$a_i\xrightarrow[i\to\infty]{}0$，所以$\sqrt{a_i}\exp\left(-b\sum_{j=i+1}^{n}a_j\right)\xrightarrow[n\to\infty]{}0$对$i\geqslant1$一致，或者$f(n,i)\triangleq\frac{1}{\sqrt{a_i}}\exp\left(b\sum_{j=i+1}^{n}a_j\right)\xrightarrow[n\to\infty]{}\infty$对$i$一致. 那么利用式(4.1.11)对任意$\varepsilon>0$就有

$$\sup E\left(\|\omega_{i+1}\|^2I_{\left[\left\|\sum_{j=0}^{\infty}C_j\right\|\|\omega_{i+1}\|>\varepsilon f(n,i)\right]}\right)\xrightarrow[n\to\infty]{}0, \tag{4.1.42}$$

所以

$$\sum_{i=1}^{n} E\left(\left\|\psi_{n,i+1}\sqrt{a_i}\left(\sum_{j=0}^{\infty}C_j\right)\omega_{i+1}\right\|^2 I_{[\|\psi_{n,i+1}\sqrt{a_i}(\sum_{j=0}^{\infty}C_j)\omega_{i+1}\|>\varepsilon]}\right)$$

$$\leqslant\sum_{i=1}^{n}a_i\|\psi_{n,i+1}\|^2\left\|\sum_{j=0}^{\infty}C_j\right\|^2 E\|\omega_{i+1}\|^2 I_{[\|\sum_{j=0}^{\infty}C_j\|\|\omega_{i+1}\|>\eta(n,i)]}.$$

由引理 4.1.2 及式(4.1.27)、(4.1.42)知，上式当 $n\to\infty$ 趋于 0.

所以据引理 4.1.1，我们知式(4.1.37)成立，同时也完成了定理的证明. □

注 4.1.1 在定理 4.1.1 的条件下，如果正整数 $n_k>\cdots>n_2>n_1>a_n^{-1}$，并且 $\lim\limits_{n\to\infty}\log(a_n n_k)=t_k$，那么可证当 $n\to\infty$ 时，$\left(\frac{1}{\sqrt{a_{n_1}}}u_{n_1},\cdots,\frac{1}{\sqrt{a_{n_k}}}u_{n_k}\right)$的联合分布收敛到$(Z(t_1),\cdots,Z(t_k))$的联合分布，这里 $Z(t)$是平稳高斯马氏过程，满足下面的随机微分方程

$$\mathrm{d}Z_t=AZ_t\mathrm{d}t+\sum_{j=0}^{\infty}C_jS_0^{\frac{1}{2}}\mathrm{d}\omega_t, \qquad (4.1.43)$$

ω_t 为 l 维标准 Wiener 过程，$S_0^{\frac{1}{2}}$ 是 S_0 的平方根阵. 证明方法和文献[74]中的定理 6.3 的证法类似.

推论 4.1.1 由式(4.1.26)和(4.1.6)，类似地可有

$$\frac{u_n}{\sqrt{a_n}}=\psi_{n-1,m}\frac{u_m}{\sqrt{a_m}}+\sum_{i=m}^{n-1}\psi_{n-1,i+1}\frac{a_i}{\sqrt{a_{i+1}}}(e_{i+1}+\nu_{i+1}),$$
$$\forall m: 1\leqslant m\leqslant n, \qquad (4.1.44)$$

由式(4.1.27)知 $\psi_{n-1,m}\frac{u_m}{\sqrt{a_m}}\xrightarrow[n\to\infty]{}0$，由式(4.1.24)、引理 4.1.2 及式(4.1.27)，我们在讨论式(4.1.28)时，已证

$$\sum_{i=m}^{n}\psi_{n,i+1}\frac{a_i}{\sqrt{a_{i+1}}}\nu_{i+1}\xrightarrow[n\to\infty]{}0, \quad \forall m: 1\leqslant m\leqslant n,$$

所以从定理 4.1.1 知，对任意 $m\geqslant1$，

$$\sum_{i=m}^{n}\psi_{n,i+1}\frac{a_i}{\sqrt{a_{i+1}}}e_{i+1}\xrightarrow[n\to\infty]{d}N(0,S). \qquad (4.1.45)$$

§4.2 变界截尾 RM 算法的渐近正态性

现在我们来讨论由式(3.1.3)、(3.1.4)给出的 x_n 的渐近正态

性，我们先引出要用的条件：

A 4.2.0 $R^l \to R^l$ 的回归函数 $h(\cdot)$ 在任一有界集上有界，并且当 $x \to x^0$ 时，

$$\|h(x) - H(x - x^0)\| \leqslant c\|x - x^0\|^{1+\beta}, \tag{4.2.1}$$

$c > 0$, $\beta \in [0, 1]$ 为常数. $H + \frac{1}{2}\alpha I$ 为稳定阵，α 的定义见下面 A 4.2.1.

A 4.2.1 $\{a_n\}$ 满足以下条件：$a_n > 0$, $a_n \to 0$, $\sum_{i=1}^{\infty} a_i = \infty$, $a_{n+1}^{-1} - a_n^{-1} \to \alpha \geqslant 0$，并对某 η: $0 < \eta < \frac{\beta}{1+\beta}$, $\sum_{i=1}^{\infty} a_i^{1+\eta} < \infty$.

A 4.2.2

$$\varepsilon_n = e_n + \nu_n, \quad e_n = \sum_{j=0}^{r} C_j \omega_{n-j}, \quad r < \infty, \tag{4.2.2}$$
$$\nu_{r+1} = o(\sqrt{a_n}) \ a.s.,$$

$\{\omega_i, \mathscr{F}_i\}$ 为鞅差序列 且满足式(4.1.9)～(4.1.11).

A 4.2.3 存在二次连续可微 $R^l \to R$ 函数 $v(\cdot)$，使对任意 $\Delta_2 > \Delta_1 > 0$,

$$\sup_{\Delta_1 \leqslant \|x - x^0\| \leqslant \Delta_2} h^{\tau}(x) v_x(x) < 0 \tag{4.2.3}$$

还设存在常数 c_0，使 $\|x^*\| < c_0$, $v(x^*) \neq \inf_{\|x\| = c_0} v(x)$, x^* 定义在式(3.1.4)中.

定理 4.2.1 设 A 4.2.0～A 4.2.3 成立，x_n 由式(3.1.3)、(3.1.4)给出，那么

$$\frac{x_n - x^0}{\sqrt{a_n}} \xrightarrow[n \to \infty]{d} N(0, S), \tag{4.2.4}$$

$$S = \int_0^{\infty} e^{\left(H + \frac{1}{2}\alpha I\right)t} \sum_{i=0}^{r} C_i S_0 \sum_{i=0}^{r} C_i^{\tau} e^{\left(H^{\tau} + \frac{1}{2}\alpha I\right)t} dt. \tag{4.2.5}$$

证明 由于 $1 - \eta > \frac{1}{1+\beta}$，所以可选 δ，使

$$1 - \eta > 2\delta > \frac{1}{1+\beta}, \tag{4.2.6}$$

那么 $2 - 2\delta > 1 + \eta$，所以从 A 4.2.1 知，$\sum_{i=1}^{\infty} a_i^{2-2\delta} < \sum_{i=1}^{\infty} a_i^{1+\eta} < \infty$. 由鞅差和收敛定理 1.2.3 知，

$$\sum_{i=1}^{\infty} a_i^{1-\delta}\omega_{i+1-j}<\infty \quad a.s., \ \forall j=0,1,\cdots,r,$$

所以

$$\sum_{i=1}^{\infty} a_i^{1-\delta}e_{i+1}=\sum_{i=1}^{\infty} a_i^{1-\delta}\sum_{j=0}^{r}C_j\omega_{i+1-j}$$
$$=\sum_{j=0}^{r}C_j\sum_{i=1}^{\infty}a_i^{1-\delta}\omega_{i+1-j}<\infty \quad a.s.,$$

即式(3.2.5)成立．又由于 $\delta<\frac{1}{2}$，所以 $\nu_{n+1}=o(\sqrt{a_n})$ 包含了 $\nu_{n+1}=o(a_n^\delta)$，同样，由于已要求 $H+\frac{1}{2}\alpha I$ 稳定，所以 $H+\delta\alpha I$ 也稳定．这样，全部 A 3.2.0～A 3.2.3 成立．据定理 3.2.1，当 $n\to\infty$ 时，

$$\|x_n-x^0\|=o(a_n^\delta) \quad a.s.,$$

从式(4.2.1)知，当 $n\to\infty$ 时，

$$\|h(x_n)-H(x-x^0)\|\leqslant c\|x_n-x^0\|^{1+\beta}$$
$$=o(a_n^{\delta(1+\beta)})<o(\sqrt{a_n}) \ a.s., \tag{4.2.7}$$

从定理 3.2.1 还知，存在取整数值的随机变量 $N<\infty$，a.s.，使 $\sigma_n=\sigma_N$，$\forall n\geqslant N$．并且对一切 $n\geqslant N$，式(3.1.4)不再截尾，也就是说，对 $n\geqslant N$，算法(3.1.4)成了 RM 算法

$$x_{n+1}=x_n+a_n(h(x_n)+e_{n+1}+\nu_{n+1}), \quad \forall n\geqslant N \tag{4.2.8}$$

或

$$x_{n+1}-x^0=(I+a_nH)(x_n-x^0)+a_n(h(x_n)-H(x_n-x^0))$$
$$+a_n(e_{n+1}+\nu_{n+1}), \ n\geqslant N. \tag{4.2.9}$$

若记 $\nu'_{n+1}=h(x_n)-H(x_n-x^0)+\nu_{n+1}$，那么由式(4.2.7)及(4.2.2)仍有 $\nu'_{n+1}=o(\sqrt{a_n})$ a.s.，式(4.2.9)就改写成

$$x_{n+1}-x^0=(I+a_nH)(x_n-x^0)+a_n(e_{n+1}+\nu'_{n+1}), \quad n\geqslant N. \tag{4.2.10}$$

由式(4.1.24)知

$$\frac{x_{n+1}-x^0}{\sqrt{a_{n+1}}}=(I+a_nA_n)\frac{x_n-x^0}{\sqrt{a_n}}+\frac{a_n}{\sqrt{a_{n+1}}}(e_{n+1}+\nu'_{n+1}), \ n\geqslant N, \tag{4.2.11}$$

$$A_n = H + \frac{1}{2}\alpha I + \frac{1}{2}a_n\alpha H + o(\sqrt{a_n})(I + a_n H)$$

$$\xrightarrow[n\to\infty]{} A \triangleq H + \frac{1}{2}\alpha I. \tag{4.2.12}$$

设 B 为 R^l 中的任一 Borel 集，那么对任一正整数 m，

$$\lim_{n\to\infty} P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B\right) = \lim_{n\to\infty}\left\{P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B, m \leqslant N\right)\right.$$

$$\left. + P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B, m > N\right)\right\}$$

$$= \lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B,\ m > N\right), \tag{4.2.13}$$

上面最后的等式因为

$$\lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B,\ m \leqslant N\right) \leqslant \lim_{m\to\infty} P(m \leqslant N) = 0.$$

那么由式(4.2.11)及(4.1.26)的符号知

$$\lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B,\ m > N\right)$$

$$= \lim_{m\to\infty}\lim_{n\to\infty} P\left(\left(\psi_{n-1,m}\frac{x_m - x^0}{\sqrt{a_m}}\right.\right.$$

$$\left.\left. + \sum_{i=m}^{n-1}\psi_{n-1,i+1}\frac{a_i}{\sqrt{a_{i+1}}}(e_{i+1} + \nu'_{i+1})\right) \in B, m > N\right), \tag{4.2.14}$$

由于

$$\lim_{m\to\infty}\lim_{n\to\infty} P\left(\left(\psi_{n-1,m}\frac{x_m - x^0}{\sqrt{a_m}}\right.\right.$$

$$\left.\left. + \sum_{i=m}^{n-1}\psi_{n-1,i+1}\frac{a_i}{\sqrt{a_{i+1}}}(e_{i+1} + \nu'_{i+1})\right) \in B, m \leqslant N\right)$$

$$\leqslant \lim_{m\to\infty} P(m \leqslant N) = 0,$$

所以从式(4.2.14)知

$$\lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{x_n - x^0}{\sqrt{a_n}} \in B,\ m > N\right)$$

$$= \lim_{n\to\infty} P\left(\psi_{n-1,m}\frac{x_m - x^0}{\sqrt{a_m}}\right.$$

$$\left. + \sum_{i=m}^{n-1}\psi_{n-1,i+1}\frac{a_i}{\sqrt{a_{i+1}}}(e_{i+1} + \nu'_{i+1}) \in B\right). \tag{4.2.15}$$

推论 4.1.1 中已指出

$$\psi_{n-1,m}\frac{x_m-x^0}{\sqrt{a_m}}\xrightarrow[n\to\infty]{}0 \text{ a.s.}, \sum_{i=m}^{n-1}\psi_{n-1,i+1}\frac{a_i}{\sqrt{a_{i+1}}}\nu'_{i+1}\xrightarrow[n\to\infty]{}0 \text{ a.s.},$$

所以从式(4.2.12)、(4.2.15)得

$$\lim_{n\to\infty}P\left(\frac{x_n-x^0}{\sqrt{a_n}}\in B\right)=\lim_{n\to\infty}P\left(\sum_{i=m}^{n}\psi_{n,i+1}\frac{a_i}{\sqrt{a_{i+1}}}e_{i+1}\in B\right),$$

再由推论 4.1.1 便得出定理结论. □

推论 4.2.1 设 D 为 $l\times l$ 阵. 我们在算法中用 a_nD 来取代 a_n, 也就是考察

$$\sigma_n=\sum_{i=0}^{n-1}I_{[\|x_i+a_iDy_{i+1}\|>M_{\sigma_i}]},\ \sigma_0=0, \tag{4.2.16}$$

$$x_{n+1}=(x_n+a_nDy_{n+1})I_{[\|x_n+a_nDy_{n+1}\|\le M_{\sigma_n}]}+x^*I_{[\|x_n+a_nDy_{n+1}\|>M_{\sigma_n}]}, \tag{4.2.17}$$

这相当于用 Dy_{i+1} 来取代 y_{i+1}, 也就是用 $Dh(\cdot)$ 来取代 $h(\cdot)$, $D\varepsilon_i$ 来取代 ε_i. 这时我们把条件 A 4.2.0 改成

A 4.2.0′ $R^l\to R^l$ 的回归函数 $h(\cdot)$ 在任一有界集上有界, 当 $x\to x^0$ 时, 式(4.2.1)成立, 并且 $DH+\frac{1}{2}\alpha I$ 稳定.

这个条件和 A 4.2.0 相比, 唯一的差别是把 $H+\frac{1}{2}\alpha I$ 稳定改成了 $DH+\frac{1}{2}\alpha I$ 稳定. 那么根据定理 4.2.1, 在条件 A 4.2.0′, A 4.2.1～A 4.2.3 下, 由算法(4.2.16)、(4.2.17)给出的 x_n 渐近正态, 即

$$\frac{x_n-x^0}{\sqrt{a_n}}\xrightarrow[n\to\infty]{d}N(0,S_D), \tag{4.2.18}$$

$$S_D=\int_0^\infty e^{\left(DH+\frac{1}{2}\alpha I\right)t}DPD^\tau e^{\left(H^\tau D^\tau+\frac{1}{2}\alpha I\right)t}\,dt, \tag{4.2.19}$$

这里简记

$$P=\sum_{i=0}^{r}C_iS_0\sum_{i=0}^{r}C_i^\tau. \tag{4.2.20}$$

§4.3 渐近有效性

在推论4.2.1中我们看到，算法中把 a_n 换成 a_nD 后，估计误差 $\frac{x_n-x^0}{\sqrt{a_n}}$ 的渐近协方差阵 S 依赖于 D. 我们希望 S 尽可能小. 下面就来讨论，当 D 取什么阵时，S 达最小，也即算法具有渐近有效性.

定理4.3.1 在 $DH+\frac{1}{2}\alpha I$ 稳定的条件下，$\alpha>0$ 使式(4.2.19)表达的 S_D 达最小的 $D=-\alpha H^{-1}$, 这时

$$\min_D S_D=\alpha H^{-1}PH^{-\tau}, \tag{4.3.1}$$

这里 $H^{-\tau}\triangleq(H^{-1})^\tau$.

证明 用 $DH+\frac{1}{2}\alpha I$ 的稳定性, 对 $\left(DH+\frac{1}{2}\alpha I\right)S_D$ 求分部积分知,

$$\begin{aligned}&\left(DH+\frac{1}{2}\alpha I\right)S_D\\&=\int_0^\infty\left(de^{\left(DH+\frac{1}{2}\alpha I\right)t}\right)DPD^\tau e^{\left(H^\tau D^\tau+\frac{1}{2}\alpha I\right)t}\\&=-DPD^\tau-S_D\left(H^\tau D^\tau+\frac{1}{2}\alpha I\right),\end{aligned}$$

这就是说, S_D 满足代数方程

$$\left(DH+\frac{1}{2}\alpha I\right)S_D+S_D\left(H^\tau D^\tau+\frac{1}{2}\alpha I\right)+DPD^\tau=0. \tag{4.3.2}$$

由于 $DH+\frac{1}{2}\alpha I$ 稳定，$\alpha>0$，所以 DH 至少非退化，所以式(4.3.2)等价于

$$\left(H+\frac{\alpha D^{-1}}{2}\right)S_DD^{-\tau}+D^{-1}S_D\left(H^\tau+\frac{\alpha D^{-\tau}}{2}\right)+P=0,$$

或者

$$\left(\frac{H}{\sqrt{\alpha}}+\sqrt{\alpha}D^{-1}\right)S_D\left(\frac{H^\tau}{\sqrt{\alpha}}+\sqrt{\alpha}D^{-\tau}\right)$$

$$-\frac{1}{\alpha}HS_DH^\tau+P=0,$$

那么

$$S_D=\alpha H^{-1}\Big[\Big(\frac{H}{\sqrt{\alpha}}+\sqrt{\alpha}D^{-1}\Big)S_D\Big(\frac{H^\tau}{\sqrt{\alpha}}+\sqrt{\alpha}D^{-\tau}\Big)+P\Big]H^{-\tau}$$
$$\geqslant\alpha H^{-1}PH^{-\tau},$$

并且当 $D=-\alpha H^{-1}$ 时达到等式. □

注 4.3.1 当 $\alpha=0$ 时, DH 稳定. 这时

$$S_D=\int_0^\infty e^{DHt}DPD^\tau e^{H^\tau D^\tau t}\mathrm{d}t.$$

若取 $D=-\varepsilon H^{-1}$, $\varepsilon>0$, 那么

$$S_D=\int_0^\infty e^{-2\varepsilon t}\varepsilon^2H^{-1}PH^{-\tau}\mathrm{d}t=\frac{\varepsilon}{2}H^{-1}PH^{-\tau}\xrightarrow[\varepsilon\to0]{}0,$$

这时 S_D 的极小值达不到.

注 4.3.2 最常用的步长因子 $a_n=\frac{a}{n}$, $a>0$, 这时 $\alpha=\frac{n+1}{a}-\frac{n}{a}=\frac{1}{a}$. 据定理 4.3.1, 最优的 $D=-\alpha H^{-1}=-\frac{H^{-1}}{a}$. 那么实际的步长因子为 $a_nD=\frac{1}{n}(-H^{-1})$, 这时 $\frac{x_n-x^0}{\sqrt{a_n}}=\frac{\sqrt{n}(x_n-x^0)}{\sqrt{a}}$ 的渐近协方差阵为 $\frac{1}{a}H^{-1}PH^{-\tau}$, 所以 $\sqrt{n}(x_n-x^0)$ 的渐近协方差阵为 $H^{-1}PH^{-\tau}$.

下面考察 $a_n=\frac{1}{n}$ 这一最常用的步长因子. $\alpha=1$, 最优的 D 为 $-H^{-1}$. 这时 $\sqrt{n}(x_n-x^0)\to N(0,\ H^{-1}PH^{-\tau})$, 渐近协方差阵达到下界 $H^{-1}PH^{-\tau}$.

由于 $h(\cdot)$ 是未知的回归函数, 所以 H^{-1} 是未知阵. 因此 $D=-H^{-1}$ 不能在实际的随机逼近算法中应用.

注意到 $H=h_x(x^0)$, 很自然地想到用量测量的差商 H_n 作为对 H 的估计, 把 $-H_n^{-1}$ 来取代 D, 也就是在式(3.1.3)、(3.1.4)中用 $x_{n+1}=x_n-\frac{H_n^{-1}}{n}y_{n+1}$ 来取代 $x_{n+1}=x_n+\frac{1}{n}y_{n+1}$. 这样的算法叫适应性随机逼近[86,91], 在相当强的条件下, 可以证明

$\sqrt{n}(x_n-x^0)$的协方差阵的极限达到下界 $H^{-1}PH^{-\tau}$.

设一个线性递推估计给出对 x^0 的估计为 x_n, 如果 $\sqrt{n}(x_n-x^0)\xrightarrow[n\to\infty]{d}N(0, S)$, $S=H^{-1}PH^{-\tau}$, 那么我们称 x_n 是对 x^0 的渐近有效估计.

下面我们不采用对 H 的估计, 而用其它办法使协方差阵达到下界 $H^{-1}PH^{-\tau}$, 并且比适应性随机逼近所用的条件有实质性减弱. 这个方法就是文献[77, 82]中提出的慢衰减增益序列加平均的方法. 后来在文献[24, 25, 95, 96]中对噪声和回归函数的条件得到进一步减弱.

在式(4.1.7)中, 当 $\alpha=0$ 时, 我们称 $\{a_n\}$ 为慢衰减序列, 典型的慢衰减序列例子为 $a_n=\frac{1}{n^{\nu}}$, $0<\nu<1$. 对算法给出的 x_n 求平均, 得到

$$\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i. \tag{4.3.3}$$

下面来研究 $\sqrt{n}(\bar{x}_n-x^0)$的渐近正态性, 将证明 $\bar{x}_n$ 是渐近有效的. 先列出要用的条件.

A4.3.0 $R^l\to R^l$ 的回归函数 $h(\cdot)$在任一有界集上有界, 并且存在稳定阵 H 及 $\gamma>0$, $\beta\in(0, 1]$, 使

$$\|h(x)-H(x-x^0)\|\leqslant c_1\|x-x^0\|^{1+\beta},$$

$$\forall x\in\{x: \|x-x^0\|\leqslant\gamma\}. \tag{4.3.4}$$

A 4.3.1 $a_n>0$, $a_n\to 0$, a_n 非增,

$$a_{n+1}^{-1}-a_n^{-1}\xrightarrow[n\to\infty]{}0, \tag{4.3.5}$$

并且对某 $\delta\in(0, \beta)$,

$$\sum_{i=1}^{\infty}\frac{a_i^{\frac{1+\delta}{2}}}{i^{\frac{1}{2}}}<\infty. \tag{4.3.6}$$

A 4.3.2
$$\varepsilon_i=e_i+\nu_i, \tag{4.3.7}$$

$$a_n\sum_{i=0}^{n}e_{i+1}\to 0,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\xrightarrow[n\to\infty]{d}N(0, P), \tag{4.3.8}$$

$$E\|e_i\|^2<\infty,\ \sum_{s>-i}^{\infty}\|Ee_ie_{i+s}^{\tau}\|\leqslant c_2, \tag{4.3.9}$$

C_2 为不依赖于 i 的常数.

$$E\|\nu_i\|^2=O(a_i^{1+\delta}), \tag{4.3.10}$$

δ 和式(4.3.6)中的 δ 相同.

A 4.3.3 存在 $R^l\to R$ 的二次连续可微函数 $v(\cdot)$，使对任意 $\Delta_2>\Delta_1>0$，$\sup\limits_{\Delta_1\leqslant\|x-x^0\|\leqslant\Delta_2} h^{\tau}(x)v_x(x)<0$. 还设存在常数 c_3，使 $\|x^*\|<c_3$，$v(x^*)\neq\inf\limits_{\|x\|=c_3} v(x)$，$x^*$ 的定义在式(3.1.4)中.

注 4.3.3 从式(4.3.6)知，

$$\sum_{i=\left[\frac{n}{2}\right]}^{n}\frac{a_i^{\frac{1+\delta}{2}}}{\sqrt{i}}=o(1),$$

由于 a_n 非增，所以

$$a_n=o(n^{-\mu}),\ \mu=\frac{1}{1+\delta}\in\left(\frac{1}{2},1\right). \tag{4.3.11}$$

注 4.3.4 如果 $e_i=\sum\limits_{j=0}^{r}C_j\omega_{i-j}$，$\{\omega_i,\mathscr{F}_i\}$ 为满足式(4.1.9)～(4.1.11)的鞅差序列，那么用引理 4.1.1 立即得出

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\xrightarrow[n\to\infty]{d}N(0,P),$$

P 由式(4.2.20)给出，也就是说式(4.3.8)中的第二个要求满足.

由对鞅差和的估计公式定理 1.2.4 知

$$\sum_{i=1}^{n}e_i=\sum_{i=r}^{n}\left(\sum_{j=0}^{r}C_j\right)\omega_i=o\left(\sqrt{n}\log^{\frac{1}{2}+\varepsilon}n\right)\quad \text{a.s.},\ \forall\varepsilon>0.$$

由式 (4.3.11) 便知式 (4.3.8) 的第一个要求也成立. 而式(4.3.9)显然成立.

定理 4.3.2 设条件 A 4.3.0～A 4.3.3 成立，x_n 由式(3.1.3)、(3.1.4)给出，记 $\bar{x}_n=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$，那么 $\bar{x}_n$ 渐近有效，即

$$\sqrt{n}(\bar{x}_n-x^0)\xrightarrow[n\to\infty]{d}N(0,S),\ S=H^{-1}PH^{-\tau}.$$

在证明定理之前，我们需要弄清慢衰减增益 a_i 的一些性质.

和式(3.2.10)一样类似地，记

$$\Phi_{n,j}=\prod_{i=j}^{n}(I+a_iH),\ n\geqslant j,\ \Phi_{j,j+1}=I, \qquad (4.3.12)$$

从式(3.2.11)知

$$\|\Phi_{n,j}\|\leqslant c_0\exp\left(-c\sum_{k=j}^{n}a_k\right),\ \ \forall n\geqslant j,\ \forall j\geqslant 0. \qquad (4.3.13)$$

记

$$G_{n,j}=\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}+H^{-1}\Phi_{n,j}. \qquad (4.3.14)$$

引理4.3.1 1)用 $o(1)$ 表示当 $j\to\infty$ 时趋于0的量，那么

$$\frac{a_j}{a_n}\leqslant\exp\left(o(1)\sum_{k=j}^{n}a_k\right),\ \ \forall n\geqslant j,\ \forall j\geqslant 1. \qquad (4.3.15)$$

2) $G_{n,j}$ 对 n 及 j 一致有界，并且

$$\frac{1}{n}\sum_{j=1}^{n}\|G_{n,j}\|\xrightarrow[n\to\infty]{}0. \qquad (4.3.16)$$

证明 从式(4.3.5)知

$$\frac{a_j}{a_{j+1}}=1+o(1)a_j,\ o(1)\xrightarrow[j\to\infty]{}0,$$

所以

$$\begin{aligned}\frac{a_j}{a_n}&=\frac{a_j}{a_{j+1}}\cdots\frac{a_{n-1}}{a_n}\\&=\prod_{k=j}^{n-1}(1+o(1)a_k)\leqslant\prod_{k=j}^{n-1}e^{o(1)a_k}\\&=\exp\left(o(1)\sum_{k=j}^{n-1}a_k\right),\end{aligned}$$

因为 $a_n\to 0$，由上式即得式(4.3.15)。

从式(4.3.11)可知，$na_n\xrightarrow[n\to\infty]{}\infty$，所以对任一 $\varepsilon>0$，我们有

$$\sum_{i=[(1-\varepsilon)n]}^{n}a_i=\sum_{i=[(1-\varepsilon)n]}^{n}\frac{ia_i}{i}\xrightarrow[n\to\infty]{}\infty. \qquad (4.3.17)$$

因此，由式(4.3.13)知

$$\begin{aligned}&\frac{1}{n}\sum_{j=0}^{n}\|\Phi_{n,j}\|\\&\leqslant\frac{1}{n}\left(\sum_{j=0}^{[(1-\varepsilon)n]}+\sum_{j=[(1-\varepsilon)n]+1}^{n}\right)c_0\exp\left(-c\sum_{k=j}^{n}a_k\right)\to 0,\end{aligned}$$

注意到式(4.3.17)，便知当 $n\to\infty$ 然后 $\varepsilon\to 0$ 时，上式趋于 0. 所以为证式(4.3.16)，只要证明 $\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}$ 对 n 及 j 一致有界，并且

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}\xrightarrow[n\to\infty]{}0. \tag{4.3.18}$$

注意到对任何 $\lambda>0$，及式(4.3.11)蕴含着 $\sum_{i=1}^{\infty}a_i^2<\infty$，我们有

$$\begin{aligned}
&\sum_{i=j}^{n}a_i\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)\\
&\quad\leqslant\frac{1}{\lambda}\sum_{i=j}^{n}\Big(1-e^{-\lambda a_i}+\frac{\lambda^2a_i^2}{2!}\Big)\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)\\
&\quad=\frac{1}{\lambda}\sum_{i=j}^{n}\Big(\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)-\exp\Big(-\lambda\sum_{k=j}^{i}a_k\Big)\Big)\\
&\qquad+\frac{\lambda}{2}\sum_{i=j}^{n}a_i^2\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)\\
&\quad\leqslant\frac{1}{\lambda}+\frac{\lambda}{2}\sum_{i=1}^{\infty}a_i^2<\infty,
\end{aligned}\tag{4.3.19}$$

所以

$$\begin{aligned}
&\sum_{i=j}^{n}a_i\sum_{k=j}^{i-1}a_k\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)\\
&\quad\leqslant\sum_{i=j}^{n}a_i\,\frac{2}{\lambda}\Big[\exp\Big(\frac{\lambda}{2}\sum_{k=j}^{i-1}a_k\Big)\Big]\exp\Big(-\lambda\sum_{k=j}^{i-1}a_k\Big)\\
&\quad=\frac{2}{\lambda}\sum_{i=j}^{n}a_i\exp\Big(-\frac{\lambda}{2}\sum_{k=j}^{i-1}a_k\Big)\\
&\quad\leqslant\frac{4}{\lambda^2}+\frac{1}{2}\sum_{i=1}^{\infty}a_i^2.
\end{aligned}\tag{4.3.20}$$

因此，由式(4.3.5)、(4.3.13)、(4.3.15)知

$$\begin{aligned}
&\Big\|\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}\Big\|\\
&\qquad\leqslant\sum_{i=j}^{n}\sum_{k=j}^{i}(a_{k-1}-a_k)\|\Phi_{i-1,j}\|\\
&\qquad\leqslant\sum_{i=j}^{n}o(a_j)\sum_{k=j}^{i-1}a_kc_0\exp\Big(-c\sum_{k=j}^{i-1}a_k\Big)
\end{aligned}$$

$$= o(1)\sum_{i=j}^{n} a_i \sum_{k=j}^{i-1} a_k \exp\left(-\frac{c}{2}\sum_{k=j}^{i-1} a_k\right)$$

$$\leqslant o(1)\left(\frac{16}{c^2}+\frac{1}{2}\sum_{i=1}^{\infty} a_i^2\right),$$

这里 $o(1)$ 表示当 $j\to\infty$ 时趋于0的量.

由此就得式(4.3.18)，同时也证明了 $\left\|\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}\right\|$ 对 n 及 j 一致有界. □

下面对式(3.1.3)、(3.1.4)定义的 x_n 给出一个表达式. 我们来定义一列过程 $\{x_n^{(i)}\}$, $i=1, 2, \cdots$ 及一列停时 $\alpha(i)$, $i=1, 2, \cdots$. 确切地说, 设

$$x_{n+1}^{(1)} = x_0 + \sum_{i=0}^{n} a_i y_{i+1}^{(1)},\ \forall n\geqslant 0,\ y_{i+1}^{(1)} = h(x_i^{(1)}) + e_{i+1}, \tag{4.3.21}$$

$$\alpha_0 = 0,\ \alpha_1 = \begin{cases}\min\{i: i\geqslant \alpha_0,\ \|x_i^{(1)}\| > M_0\},\\ \infty, \text{如果} \sup_i \|x_i^{(1)}\| \leqslant M_0.\end{cases} \tag{4.3.22}$$

显然
$$y_{i+1}^{(1)} = y_{i+1},\ \forall i < \alpha_1.$$
然后定义

$$\begin{aligned} x_{n+1}^{(2)} &= x^* + \sum_{i=0}^{n} a_i y_{i+1}^{(2)} I_{[\alpha_1 \leqslant i]},\ \forall n\geqslant 0,\\ y_{i+1}^{(2)} &= h(x_i^{(2)}) + e_{i+1},\ i\in[\alpha_1, \alpha_2). \end{aligned} \tag{4.3.23}$$

$$\alpha_2 = \begin{cases}\min\{i: i> \alpha_1,\ \|x_i^{(2)}\| > M_1\},\\ \infty, \text{如果} \sup_i \|x_i^{(2)}\| \leqslant M_1.\end{cases} \tag{4.3.24}$$

递推地定义

$$\begin{aligned} x_{n+1}^{(k)} &= x^* + \sum_{i=0}^{n} a_i y_{i+1}^{(k)} I_{[\alpha_{k-1} \leqslant i]},\ \forall n\geqslant 0,\\ y_{i+1}^{(k)} &= h(x_i^{(k)}) + e_{i+1},\ i\in[\alpha_{k-1}, \alpha_k). \end{aligned} \tag{4.3.25}$$

$$\alpha_k = \begin{cases}\min\{i: i> \alpha_{k-1},\ \|x_i^{(k)}\| > M_{k-1}\},\\ \infty, \text{如果} \sup_i \|x_i^{(k)}\| \leqslant M_{k-1}.\end{cases} \tag{4.3.26}$$

实际上, α_k 是 α_{k-1} 以后 $x_i^{(k)}$ 首次逸出半径为 M_{k-1} 的球的时间, 而在时间区间 $[\alpha_{k-1}, \alpha_k)$ 内, $x_n^{(k)}$ 和式(3.1.3)、(3.1.4)定义的 x_n 相

间，也和 RM 过程的递推关系相同．所以由式(3.1.3)、(3.1.4)定义的 x_n 可表达为

$$x_n=\sum_{i=1}^{\infty} x_n^{(i)} I_{[\alpha_{i-1}<n<\alpha_i]}. \tag{4.3.27}$$

根据定理 3.1.3，$\{x_n\}$ a. s. 有界，因此存在取整数值的 i_n（依赖于样本），使 $\alpha_{i0}<\infty$ a. s.，$\alpha_{i_0+1}=\infty$ a. s.，所以

$$x_{n+1}=x_n+a_n y_{n+1},\ \forall n\geqslant\alpha_{i_0}, \tag{4.3.28}$$

因此

$$x_{n+1}-x^0=(I+a_n H)(x_n-x^0)+a_n(h(x_n)-H(x_n-x^0))$$
$$+a_n\varepsilon_{n+1},\ n\geqslant\alpha_{i_0}, \tag{4.3.29}$$

或者用式(4.3.12)知

$$x_{n+1}-x^0=\Phi_{n,\alpha_{i_0}}(x_{\alpha_{i_0}}-x^0)+\sum_{j=\alpha_{i_0}}^{n}\Phi_{n,j+1}a_j\varepsilon_{j+1}$$
$$+\sum_{j=\alpha_{i_0}}^{n}\Phi_{n,j+1}a_j(h(x_j)-H(x_j-x^0)),\ \forall n\geqslant\alpha_{i_0}. \tag{4.3.30}$$

对式(4.3.4)中固定的 γ 以及确定性的 n_0 定义停时 μ:

$$\mu=\begin{cases}\min\{j: j>n_0,\ \|x_j-x^0\|\geqslant\gamma\},\\ 0,\ 如果\|x_{n_0}-x^0\|\geqslant\gamma,\end{cases} \tag{4.3.31}$$

在 $[\alpha_{i_0}<n_0]$ 上，从式(4.3.29)中可得

$$(x_{n+1}-x^0)I_{[\alpha_{i_0}<n_0]}$$
$$=\Phi_{n,n_0}(x_{n_0}-x^0)I_{[\alpha_{i_0}<n_0]}+\sum_{j=n_0}^{n}\Phi_{n,j+1}a_j\varepsilon_{j+1}I_{[\alpha_{i_0}<n_0]}$$
$$+\sum_{j=n_0}^{n}\Phi_{n,j+1}a_j(h(x_j)-H(x_j-x^0))I_{[\alpha_{i_0}<n_0]},\ \forall n\geqslant n_0.$$
$$\tag{4.3.32}$$

由此便知在 $[\alpha_{i_0}<n_0]$ 上，估计误差在球 $\{\|x_i-x^0\|<\gamma\}$ 内的表达式为

$$(x_{n+1}-x^0)I_{[\mu>n+1,\alpha_{i_0}<n_0]}$$
$$=\Phi_{n,n_0}(x_{n_0}-x^0)I_{[\mu>n+1,\alpha_{i_0}<n_0]}$$
$$+\sum_{j=n_0}^{n}\Phi_{n,j+1}a_j\varepsilon_{j+1}I_{[\mu>n+1,\alpha_{i_0}<n_0]}$$

$$+\sum_{j=n_0}^{n}\Phi_{n,j+1}a_j(h(x_j)-H(x_j-x^0))I_{[\mu>j,a_{i_0}<n_0]}$$

$$\cdot I_{[\mu>n+1,a_{i_0}<n_0]},\quad \forall n\geqslant n_0. \tag{4.3.33}$$

引理 4.3.2　在定理 4.3.2 的条件下,

$$\frac{1}{a_{n+1}}\mathrm{E}\|(x_{n+1}-x^0)I_{[\mu>n+1,a_{i_0}<n_0]}\|^2$$

对 n 一致有界.

证明　从式(4.3.33),利用式(4.3.13)及条件(4.3.4),我们有

$$\begin{aligned}
&\frac{1}{a_{n+1}}\mathrm{E}\Big(\|x_{n+1}-x^0\|^2I_{[\mu>n+1,a_{i_0}<n_0]}\\
&\quad\leqslant\frac{4c_0^2r^2}{a_{n+1}}\exp\Big(-2c\sum_{k=n_0}^{n}a_k\Big)\\
&\quad+\frac{4c_0^2}{a_{n+1}}\sum_{i=n_0}^{n}\sum_{j=n_0}^{n}\Big[\exp\Big(-c\sum_{k=j+1}^{n}a_k\Big)\Big]\\
&\quad\cdot a_j\Big[\exp\Big(-c\sum_{k=i+1}^{n}a_k\Big)\Big]a_i\|\mathrm{E}e_{i+1}e_{j+1}^{\tau}\|\\
&\quad+\frac{4c_0^2}{a_{n+1}}\sum_{i=n_0}^{n}\sum_{j=n_0}^{n}\Big[\exp\Big(-c\sum_{k=j+1}^{n}a_k\Big)\Big]\\
&\quad\cdot a_j\Big[\exp\Big(-c\sum_{k=i+1}^{n}a_k\Big)\Big]a_i\mathrm{E}\|\nu_{i+1}\nu_{j+1}^{\tau}\|\\
&\quad+\frac{4}{a_{n+1}}\mathrm{E}\Big(\sum_{j=n_0}^{n}c_0c_1\Big[\exp\Big(-c\sum_{k=j+1}^{n}a_k\Big)\Big]\\
&\quad\cdot a_j\|x_j-x^0\|^{1+\delta}I_{[\mu>j,a_{i_0}<n_0]}\Big)^2\\
&=I_1+I_2+I_3+I_4,
\end{aligned}\tag{4.3.34}$$

这里 I_i, $i=1, 2, 3, 4$ 分别表示式(4.3.34)右端诸式.

由式(4.3.15)知

$$I_1\leqslant\frac{4c_0^2r^2}{a_{n_0}}\exp\Big(o(1)\sum_{k=n_0}^{n+1}a_k-2c\sum_{k=n_0}^{n}a_k\Big),$$

这里 $o(1)\xrightarrow[n_0\to\infty]{}0$,所以当 n_0 充分大时,I_1 对 n 一致有界.

由式(4.3.9)、(4.3.15)知当 n_0 充分大时,

$$I_2\leqslant 4c_0^2\sum_{i=n_0}^{n}\sum_{j=n_0}^{n-i}\left[\exp\left(-\frac{c}{2}\sum_{k=j+1}^{n}a_k\right)\right]$$

$$\cdot a_i\left[\exp\left(-c\sum_{k=i+1}^{n}a_k\right)\right]\|\mathrm{E}e_{i+1}e_{j+1}^{\tau}\|$$

$$\leqslant 4c_0^2\sum_{i=n_0}^{n}\sum_{s=n_0-i}^{n-i}a_i\|\mathrm{E}\,e_{i+1}e_{i+s+1}^{\tau}\|\ \exp\left(-c\sum_{k=i+1}^{n}a_k\right)$$

$$\leqslant 4c_0^2c_2\sum_{i=n_0}^{n}a_i\exp\left(-c\sum_{k=i+1}^{n}a_k\right),\tag{4.3.35}$$

在式(4.1.16)的推导中，只要把 cr 换成 c，便知上式对 n 一致有界.

由式(4.3.10)、(4.3.15)知，当 n_0 充分大时，

$$I_3\leqslant 4c_0\sum_{i=n_0}^{n}\sum_{j=n_0}^{n}\left[\exp\left(-\frac{c}{2}\sum_{k=j+1}^{n}a_k\right)\right]a_j^{\frac{1}{2}}$$

$$\cdot\left[\exp\left(-\frac{c}{2}\sum_{k=i+1}^{n}a_k\right)\right]a_i^{\frac{1}{2}}O\left(a_i^{\frac{1+\delta}{2}}a_j^{\frac{1+\delta}{2}}\right)$$

$$\leqslant 4c_4\sum_{i=n_0}^{n}a_i\left[\exp\left(-\frac{c}{2}\sum_{k=i+1}^{n}a_k\right)\right]$$

$$\cdot\sum_{j=n_0}^{n}a_j\exp\left(-\frac{c}{2}\sum_{k=j+1}^{n}a_k\right),\ c_4\text{ 为常数},\tag{4.3.36}$$

与式(4.1.16)一样，类似地知上式对 n 一致有界.

最后，我们对 I_4 估计如下.

$$I_4\leqslant\frac{4c_0^2c_1^2}{a_{n+1}}\sum_{j=n_0}^{n}a_j^2\exp\left(-c\sum_{k=j+1}^{n}a_k\right)$$

$$\cdot\sum_{j=n_0}^{n}\left[\exp\left(-c\sum_{k=j+1}^{n}a_k\right)\right]\mathrm{E}\left(\|x_j-x^0\|^{2(1+\delta)}I_{[\mu>j,\alpha_{i_0}<n_0]}\right),$$

而当 n_0 充分大时，

$$\sum_{j=n_0}^{n}\frac{a_j^2}{a_{n+1}}\exp\left(-c\sum_{k=j+1}^{n}a_k\right)$$

$$\leqslant\sum_{j=n_0}^{n}a_j\exp\left(-\frac{c}{2}\sum_{k=j+1}^{n}a_k\right),$$

而如上所述，它对 n 一致有界，所以存在常数 c_5，使

$$I_4\leqslant c_5\gamma^{2\delta}\sum_{j=n_0}^{n}\exp\left(-c\sum_{k=j+1}^{n}a_k\right)a_j\mathrm{E}\left(\frac{\|x_j-x^0\|^2}{a_j}I_{[\mu>j,\alpha_{i_0}<n_0]}\right).\tag{4.3.37}$$

综合(4.3.34)～(4.3.37)知存在常数 k_0，使

$$\frac{1}{a_{n+1}}\mathrm{E}(\|x_{n+1}-x^0\|^2 I_{[\mu>n+1,a_{i_0}<n_0]})$$
$$\leqslant k_0+c_5\gamma^{2\beta}\sum_{j=n_0}^{n}a_j\left[\exp\left(-c\sum_{k=j+1}^{n}a_k\right)\right]$$
$$\cdot\mathrm{E}\left(\frac{\|x_j-x^0\|^2}{a_j}I_{[\mu>j,a_{i_0}<n]}\right). \tag{4.3.38}$$

记

$$g_n=\left[\exp\left(c\sum_{k=1}^{n}a_k\right)\right]\frac{1}{a_n}\mathrm{E}(\|x_n-x^0\|^2 I_{[\mu>n,a_{i_0}<n_0]}), \tag{4.3.39}$$

$$p_{n+1}=k_0\exp\left(c\sum_{k=1}^{n+1}a_k\right), \tag{4.3.40}$$

那么，从式(4.3.38)知

$$g_{n+1}\leqslant p_{n+1}+c_5\gamma^{2\beta}\sum_{j=n_0}^{n}a_j e^{ca_{n+1}}g_j$$
$$\leqslant p_{n+1}+k_1\gamma^{2\beta}\sum_{j=n_0}^{n}a_j g_j, \tag{4.3.41}$$

k_1 为常数.

记
$$f_n=\sum_{j=n_0}^{n}a_j g_j,$$

那么，从式(4.3.41)得

$$f_{n+1}=a_{n+1}g_{n+1}+f_n\leqslant(1+k_1\gamma^{2\beta}a_{n+1})f_n+a_{n+1}p_{n+1}$$
$$\leqslant\cdots\leqslant\sum_{i=n_0}^{n+1}\prod_{j=i+1}^{n+1}(1+k_1\gamma^{2\beta}a_j)a_i p_i, \tag{4.3.42}$$

这里，当 $i=n+1$ 时定义 $\prod_{j=i+1}^{n+1}(1+k_1\gamma^{2\beta}a_j)=1$.

从式(4.3.41)、(4.3.42)知

$$g_{n+1}\leqslant p_{n+1}+k_1\gamma^{2\beta}\sum_{i=n_0}^{n}\prod_{j=i+1}^{n}\left[\exp\left(k_1 2\beta\sum_{j=i+1}^{n}a_j\right)\right]a_i p_i.$$

把 γ 取得充分小，使

$$-c+k_1\gamma^{2\beta}\triangleq-k_2<0,$$

那么，从式(4.3.39)、(4.3.40)及(4.3.42)得

$$\frac{1}{a_{n+1}}\mathrm{E}\|(x_{n+1}-x^0)I_{[\mu>n+1,\alpha_{i_0}<n_0]}\|^2$$

$$\leqslant k_0+k_0k_1\gamma^{2\beta}\sum_{i=n_0}^{n}a_i\exp\left(-k_2\sum_{j=i+1}^{n}a_j\right), \tag{4.3.43}$$

而上式右端对 n 一致有界，已在式(4.1.16)、(4.3.35)等中多次证明. □

引理 4.3.3 在定理 4.3.2 的条件下,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\|h(x_i)-H(x_i-x^0)\|\xrightarrow[n\to\infty]{}0 \quad \text{a.s.}. \tag{4.3.44}$$

证明 用 Kronecker 引理 1.2.6, 我们只要证明, 对充分大的 n_0,

$$\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|<\infty \quad \text{a.s.}, \tag{4.3.45}$$

用引理 4.3.2 知存在常数 c_6 使

$$\mathrm{E}\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|I_{[\mu>i,\alpha_{i_0}<n_0]}$$

$$\leqslant \mathrm{E}\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}c_1\|x_i-x^0\|^{1+\beta}I_{[\mu>i,\alpha_{i_0}<n_0]}$$

$$\leqslant c_6\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}a_i^{\frac{1+\beta}{2}},$$

由式(4.3.11)并注意到 $\beta>\delta$, 从上式知

$$\mathrm{E}\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|I_{[\mu>i,\alpha_{i_0}<n_0]}$$

$$<c_6\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\left(i^{-\frac{(1-\delta)}{2(1+\delta)}}\right)<\infty,$$

所以 $\displaystyle\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|I_{[\mu>i,\alpha_{i_0}<n_0]}<\infty$, a.s.,

这表明

$$\left\{\sum_{i=n_0}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|<\infty\right\}$$

$$\supset\{\sup_{n_0\leqslant i<\infty}\|x_i-x^0\|<\gamma,\ \alpha_{i_0}<n_0\}. \tag{4.3.46}$$

根据定理 3.1.3, $\|x_n-x^0\|\xrightarrow[n\to\infty]{}0$ a.s., 另外 $\alpha_{i_0}<\infty$ a.s.. 所以对任一 $\varepsilon>0$, 只要 n_0 充分大, 就有

$$P\{\sup_{n_i\le i<\infty}\|x_i-x^0\|<\gamma,\ \alpha_{i_0}<n_0\}>1-\varepsilon,$$

这和式(4.3.46)一起表明

$$P\left\{\sum_{i=n_i}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|<\infty\right\}>1-\varepsilon,$$

或等价地 $P\left\{\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}\|h(x_i)-H(x_i-x^0)\|<\infty\right\}>1-\varepsilon.$

由于 ε 可任意小，这就证明了式(4.3.45). □

定理4.3.2的证明

根据定理3.1.3，算法(3.1.3)、(3.1.4)对几乎所有样本，截尾次数有穷，即 $\sigma_n\xrightarrow[n\to\infty]{}N<\infty$, a.s.，所以

$$\begin{aligned}
\sqrt{n}\,(\bar{x}_n-x^0)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-x^0)\\
&=o(1)+\frac{1}{\sqrt{n}}\sum_{i=N}^{n}(x_i-x^0)\\
&=\frac{1}{\sqrt{n}}\sum_{i=N}^{n}\Phi_{i-1,N}(x_N-x^0)\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=N}^{n}\sum_{j=N}^{i-1}\Phi_{i-1,j+1}a_j\varepsilon_{j+1}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=N}^{n}\sum_{j=N}^{i-1}\Phi_{i-1,j+1}a_j[h(x_j)\\
&\quad-H(x_j-x^0)]+o(1).
\end{aligned}\tag{4.3.47}$$

由于 $$\Phi_{n,j}=\Phi_{n-1,j}+a_nH\Phi_{n-1,j},$$

所以 $\Phi_{n,j}=I+\sum_{i=j}^{n}a_iH\phi_{i-1,j},\ H^{-1}\Phi_{n,j}=H^{-1}+\sum_{i=j}^{n}a_i\Phi_{i-1,j},$

因此

$$\begin{aligned}
a_{j-1}\sum_{i=j}^{n}\Phi_{i-1,j}&=\sum_{i=j}^{n}(a_{j-1}-a_i)\Phi_{i-1,j}\\
&\quad+\sum_{i=j}^{n}a_i\Phi_{i-1,j}=-H^{-1}+G_{n,j},
\end{aligned}\tag{4.3.48}$$

由此及 $G_{n,j}$ 的有界性(见引理4.3.1)知

$$\frac{1}{\sqrt{n}}\sum_{i=N}^{n}\Phi_{i-1,N}(x_N-x^0)$$

$$=\frac{1}{a_{N-1}\sqrt{n}}(-H^{-1}+G_{n,N})(x_N-x^0)\xrightarrow[n\to\infty]{}0 \quad \text{a.s.}, \tag{4.3.49}$$

从式(4.3.48)及 $G_{n,j}$ 的有界性，我们知 $\sum_{i=j+1}^{n}\Phi_{i-1,j+1}a_j$ 对 n 及 j 一致有界，用 c_7 表示这个常数，那么由式(4.3.44)知

$$\frac{1}{\sqrt{n}}\left\|\sum_{i=N}^{n}\sum_{j=N}^{i-1}\Phi_{i-1,j+1}a_j[h(x_j)+H(x_j-x^0)]\right\|$$

$$=\frac{1}{\sqrt{n}}\left\|\sum_{j=N}^{n}\sum_{i=j+1}^{n}\Phi_{i-1,j+1}a_j[h(x_j)-H(x_j-x^0)]\right\|$$

$$\leqslant\frac{c_7}{\sqrt{n}}\sum_{j=N}^{n}\|h(x_j)-H(x_j-x^0)\|\xrightarrow[n\to\infty]{}0 \text{ a.s.}. \tag{4.3.50}$$

把式(4.3.49)、(4.3.50)代入式(4.3.47)，并由式(4.3.48)便知

$$\sqrt{n}(\bar{x}_n-x^0)=\frac{1}{\sqrt{n}}\sum_{j=N}^{n}\sum_{i=j+1}^{n}\Phi_{i-1,j+1}a_je_{j+1}+o(1)$$

$$=\frac{-H^{-1}}{\sqrt{n}}\sum_{j=N}^{n}e_{j+1}+\frac{1}{\sqrt{n}}\sum_{j=N}^{n}G_{n,j+1}e_{j+1}$$

$$+\frac{1}{\sqrt{n}}\sum_{j=N}^{n}(-H^{-1}+G_{n,j+1})\nu_{j+1}+o(1). \tag{4.3.51}$$

由引理4.3.1之2)及式(4.3.9)知

$$\mathrm{E}\left\|\frac{1}{\sqrt{n}}\sum_{j=0}^{n}G_{n,j+1}e_{j+1}\right\|^2$$

$$\leqslant\frac{1}{n}\sum_{j=0}^{n}\|G_{n,j+1}\|\sum_{s=-j}^{n-j}\|G_{n,j+s}\|\|\mathrm{E}e_{j+s+1}e_{j+1}^{\tau}\|$$

$$\leqslant\frac{c_2c_8}{n}\sum_{j=0}^{n}\|G_{n,j+1}\|\xrightarrow[n\to\infty]{}0, \tag{4.3.52}$$

这里 c_8 表示 $\|G_{n,j}\|$ 的上界。

从式(4.3.52)知

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n}G_{n,j+1}e_{j+1}\xrightarrow[n\to\infty]{P}0,$$

再由 $G_{n,j}$ 的有界性，知

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{N-1}G_{n,j+1}e_{j+1}\xrightarrow[n\to\infty]{}0 \text{ a.s.},$$

所以

$$\frac{1}{\sqrt{n}}\sum_{j=N}^{n} G_{n,j+1} e_{j+1} \xrightarrow[n\to\infty]{p} 0. \qquad (4.3.53)$$

注意到式(4.3.6)就有

$$\mathrm{E}\sum_{j=1}^{n}\frac{1}{\sqrt{j}}\mathrm{E}\|\nu_{j+1}\| \leqslant \sum_{j=1}^{n}\frac{1}{\sqrt{j}}\,O(a_j^{\frac{1+\delta}{2}})<\infty,$$

所以用 Kronecker 引理知

$$\mathrm{E}\,\frac{1}{\sqrt{n}}\left\|\sum_{j=N}^{n}(-H^{-1}+G_{n,j+1})\nu_{j+1}\right\|$$

$$\leqslant(\|H^{-1}\|+c_8)\,\frac{1}{\sqrt{n}}\sum_{j=0}^{n}\mathrm{E}\|\nu_{j+1}\| \xrightarrow[n\to\infty]{} 0.$$

因此

$$\frac{1}{\sqrt{n}}\sum_{j=N}^{n}(-H^{-1}+G_{n,j+1})\nu_{j+1} \xrightarrow[n\to\infty]{p} 0. \qquad (4.3.54)$$

把式(4.3.53)、(4.3.54)代入到式(4.3.51)，便知

$$\sqrt{n}\,(\bar{x}_n-x^0)+\frac{H^{-1}}{\sqrt{n}}\sum_{j=N}^{n} e_{j+1} \xrightarrow[n\to\infty]{p} 0, \qquad (4.3.55)$$

由于 $\frac{1}{\sqrt{n}}\sum_{j=0}^{N-1} e_{j+1} \xrightarrow[n\to\infty]{} 0$，所以从式(4.3.55)及式(4.3.8)中便得出定理结论。 □

第 5 章

连续时间的随机逼近

我们仍设 $h(\cdot)$ 是 $R^l \to R^l$ 的回归函数，要求它的零点 x^0：$h(x^0)=0$. 和前几章不同，现在讨论量测是连续时间过程的情形

$$y_t=\int_{t_0}^t h(x_s)\mathrm{d}s+\varepsilon_t+y_{t_0},$$

ε_t 是量测误差，除了 Wiener 过程驱动的噪声外，它还包含其它"小"噪声. 对连续时间的随机逼近，首先要保证随机微分方程解的存在，然后才能分析它的收敛性、渐近正态性及有效性，而对离散时间，解的存在性不成问题. 本章的主要参考文献为[18～21, 25, 29].

§5.1 连续时间的 RM 算法

设连续时间的量测为

$$y_t=\int_{t_0}^t h(x_s)\mathrm{d}s+\varepsilon_t+y_{t_0}, \tag{5.1.1}$$

$$\varepsilon_t=\int_{t_0}^t F(x_s)\mathrm{d}w_s,$$

$(w_t, \mathscr{F}_t)$ 是 Wiener 过程，$\mathscr{F}_t\uparrow$.

设 $\{a_t\}$ 是增益函数，那么当下列随机微分方程

$$x_t=x_{t_0}+\int_{t_0}^t a_s\mathrm{d}y_s \tag{5.1.2}$$

或

$$\mathrm{d}x_t=a_t h(x_t)\mathrm{d}t+a_t F(x_t)\mathrm{d}w_t \tag{5.1.3}$$

有强解 x_t 时，则 x_t 当作对 x^0 的估计，而式(5.1.2)或(5.1.3)就是

连续时间的 RM 算法。

我们要用下列条件.

A 5.1.0 存在常数 $c_0>0$,使对任意 x, $y\in R^l$ 有

$$\|h(x)-h(y)\|+\|F(x)-F(y)\|\leqslant c_0\|x-y\|. \qquad (5.1.4)$$

A 5.1.1 $a_t>0$,

$$\int_{t_0}^{\infty} a_t\,\mathrm{d}t=\infty, \quad \int_{t_0}^{\infty} a_t^2\,\mathrm{d}t<\infty. \qquad (5.1.5)$$

A 5.1.2 $\{\mathscr{F}_t\}$ 是非降右连续 σ-代数流, $(w_t, \mathscr{F}_t)$ 是 l-维 Wiener 过程.

A 5.1.3 对任意 $\varepsilon>0$,

$$\sup_{\|x-x^0\|\geqslant\varepsilon} h(x)(x-x^0)=-\alpha_\varepsilon<0. \qquad (5.1.6)$$

定理 5.1.1 设A5.1.0～A5.1.3成立,那么连续时间的 RM 算法对任意初值 x_{t_0} 有唯一强解 x_t,并且

$$x_t \xrightarrow[t\to\infty]{} x^0 \quad \text{a.s.}.$$

证明 首先条件(5.1.4)保证了存在唯一强解(定理1.3.2).

对任意 $\varepsilon>0$,用 τ_ε 表示 x_t 初出

$$G_\varepsilon=\{x:\ \|x_t-x^0\|\geqslant\varepsilon\} \qquad (5.1.7)$$

的时间,即

$$\tau_\varepsilon=\inf\{t\geqslant 0:\ \|x_t-x^0\|<\varepsilon\}. \qquad (5.1.8)$$

根据引理1.2.3, τ_ε 为停时.

对式(5.1.3)用 $I_{t\hat{o}}$ 公式有

$$\mathrm{d}\|x_t-x^0\|^2=2(x_t-x^0)^\tau(a_t h(x_t)\mathrm{d}t+a_t F(x_t)\mathrm{d}w_t) + a_t^2\,\mathrm{tr}\,F(x_t)F^\tau(x_t)\mathrm{d}t. \qquad (5.1.9)$$

从式(5.1.4)容易看出 $\|F(x)\|^2\leqslant c_0\|x\|^2$, 由此知存在常数 $c>0$,使

$$\mathrm{tr}\,F(x)F^\tau(x)\leqslant l\|F(x)\|^2\leqslant c(1+\|x-x^0\|^2). \qquad (5.1.10)$$

记

$$v_t(x)=(1+\|x-x^0\|^2)\exp\left[c\int_t^\infty a_s^2\,\mathrm{d}s\right]. \qquad (5.1.11)$$

用 Ito 公式及(5.1.9)知

$$\mathrm{d}v_t(x_t)=-ca_t^2v_t(x_t)\mathrm{d}t+\left[\exp\left(c\int_t^\infty a_s^2\,\mathrm{d}s\right)\right]$$
$$\cdot[2(x_t-x^0)^\tau(a_th(x_t)\mathrm{d}t+a_tF(x_t)\mathrm{d}w_t)$$
$$+a_t^2\mathrm{tr}F(x_t)F^\tau(x_t)\mathrm{d}t],$$

对 $t\geqslant s$ 有

$$\mathrm{E}(v_t(x_t)|\mathscr{F}_s)$$
$$=v_s(x_s)+\mathrm{E}\left\{\left[-\int_s^t ca_\lambda^2v_\lambda(x_\lambda)\,\mathrm{d}\lambda+2\int_s^t\left(\exp\left(c\int_\lambda^\infty a_\mu^2\,\mathrm{d}\mu\right)\right)\right.\right.$$
$$\cdot(x_\lambda-x^0)^\tau a_\lambda h(x_\lambda)\mathrm{d}\lambda$$
$$\left.\left.+\int_s^t a_\lambda^2\mathrm{tr}\,F(x_\lambda)F^\tau(x_\lambda)\exp\left(c\int_\lambda^\infty a_\mu^2\,\mathrm{d}\mu\right)\mathrm{d}\lambda\right]\middle|\mathscr{F}_s\right\}.\tag{5.1.12}$$

由式(5.1.6)知

$$(x_\lambda-x^0)^\tau h(x_\lambda)<0,\quad \forall x_\lambda\neq x^0.\tag{5.1.13}$$

把式(5.1.10)代入式(5.1.12),并由式(5.1.13)知

$$\mathrm{E}(v_t(x_t)|\mathscr{F}_s)$$
$$\leqslant v_s(x_s)+\mathrm{E}\left\{\left[-\int_s^t Ca_\lambda^2v_\lambda(x_\lambda)\mathrm{d}\lambda\right.\right.$$
$$+2\int_s^t\left[\exp\left(c\int_\lambda^\infty a_\mu^2\,\mathrm{d}\mu\right)\right]a_\lambda(x_\lambda-x^0)^\tau h(x_\lambda)I_{[\tau_s>\lambda]}\,\mathrm{d}\lambda$$
$$\left.\left.+c\int_s^t a_\lambda^2v_\lambda(x_\lambda)\,\mathrm{d}\lambda\right]|\mathscr{F}_s\right\}$$
$$\leqslant v_s(x_s)-\mathrm{E}\left\{2\alpha_s\int_s^t\left[\exp\left(c\int_\lambda^\infty a_\mu^2\mathrm{d}\mu\right)\right]a_\lambda I_{[\tau_s>\lambda]}\mathrm{d}\lambda\,|\mathscr{F}_s\right\}.\tag{5.1.14}$$

记

$$b_\lambda=2\alpha_s a_\lambda\exp\left[c\int_\lambda^\infty a_\mu^2\mathrm{d}\mu\right].$$

那么由式(5.1.5)知

$$\int_{t_0}^\infty b_\lambda\,\mathrm{d}\lambda\geqslant\int_{t_0}^\infty 2\alpha_s a_\lambda\,\mathrm{d}\lambda=\infty.\tag{5.1.15}$$

从式(5.1.14)知

$$\mathrm{E}(v_t(x_t)|\mathscr{F}_s)\leqslant v_s(x_s)-\mathrm{E}\left(\int_s^t b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda|\mathscr{F}_s\right), \tag{5.1.16}$$

记

$$u_t(x)=v_t(x)I_{[\tau_\varepsilon>t]}+\int_{t_0}^t b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda. \tag{5.1.17}$$

从式(5.1.16)易见，对 $t\geqslant s\geqslant t_0$

$$\begin{aligned}
&\mathrm{E}(u_t(x_t)|\mathscr{F}_s)\\
&\leqslant E(v_t(x_t)|\mathscr{F}_s)I_{[\tau_\varepsilon>s]}+\mathrm{E}\left(\int_s^t b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda|\mathscr{F}_s\right)\\
&\quad+\int_{t_0}^s b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda\\
&\leqslant v_s(x_s)I_{[\tau_\varepsilon>s]}-\mathrm{E}\left(\int_s^t b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda|\mathscr{F}_s\right)I_{[\tau_\varepsilon>s]}\\
&\quad+\mathrm{E}\left(\int_s^t b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda|\mathscr{F}_s\right)+\int_{t_0}^s b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda\\
&=u_s(x_s).
\end{aligned} \tag{5.1.18}$$

所以$(u_t(x_t),\mathscr{F}_t)$是非负上鞅，根据鞅收敛定理知，当 $t\to\infty$ 时，$u_t(x_t)$a.s. 收敛到有穷极限，由于 $v_t(x)\geqslant 0$，从式(5.1.16)知

$$\int_{t_0}^\infty b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda<\infty \quad \text{a. s.}. \tag{5.1.19}$$

从式(5.1.15)知

$$[\tau_\varepsilon=\infty]\subset\left[\int_{t_0}^\infty b_\lambda I_{[\tau_\varepsilon>\lambda]}\mathrm{d}\lambda=\infty\right]. \tag{5.1.20}$$

由此及式(5.1.19)知 $P[\tau_\varepsilon=\infty]=0$，或

$$P[\tau_\varepsilon<\infty]=1.$$

这说明，x_t 一定进入球 $\{x:\|x-x^0\|<\varepsilon\}$. 由于 ε 任意，所以存在子序列 x_{t_i} 使$\|x_{t_i}-x^0\|\xrightarrow[i\to\infty]{}0$. 但从式(5.1.16)及上鞅收敛定理知当 $t\to\infty$ 时，$v_t(x_t)$收敛到有穷极限. 由于 $\int_{t_0}^\infty a_s^2\mathrm{d}s<\infty$，从式(5.1.11)知 $\|x_t-x^0\|^2$ 收敛到有穷极限. 但它的子列收敛到零，所以必有 $\|x_t-x^0\|\xrightarrow[t\to\infty]{}0$ a. s. □

定理 5.1.1 的证明方法简洁，但对回归函数 $h(\cdot)$ 所加的条件较强，特别是条件(5.1.4)，不仅对光滑性提出了要求，而且还限制了 $\|h(x)\|$ 的增长速度，不超过线性.

§5.2 变界截尾随机逼近算法

这一节我们来研究较一般的回归函数.

如果对任一 $n\geqslant 1$ 及任意 x、$y\in R^l$ 下式成立，

$$(\|f(x)-f(y)\|^2 I_{[\|x\|<n,\|y\|<n]})\leqslant L_n\|x-y\|^2,$$

L_n 为依赖于 n 的常数，那么称 $R^m\to R^m$ 函数 $f(\cdot)$ 满足局部 Lipschitz 条件.

设量测为

$$\begin{aligned} y_t&=y_{t_0}+\int_{t_0}^t h(x_s)\mathrm{d}s+\varepsilon_t,\\ \varepsilon_t&=\int_{t_0}^t F(x_s)\,\mathrm{d}w_s+\int_{t_0}^t \eta_s\,\mathrm{d}s. \end{aligned}\tag{5.2.1}$$

对增益函数 a_t 定义 $s(t,T)$, $\forall t\geqslant t_0$, $T>0$,

$$\int_t^{s(t,T)} a_s\,\mathrm{d}s=T.\tag{5.2.2}$$

对回归函数及量测误差 ε_t 及在算法中要用到的增益函数 a_t，引进以下条件:

A 5.2.0 $\|F(\cdot)\|$ 有界，$h(\cdot)$ 及 $F(\cdot)$ 满足局部 Lipschitz 条件，$h(x)=0$, $\forall x\in J$.

A 5.2.1 $a_t>0$,

$$\int_{t_0}^\infty a_t\,\mathrm{d}t=\infty,\qquad \int_{t_0}^\infty a_t^2\,\mathrm{d}t<\infty.\tag{5.2.3}$$

A 5.2.2 a) 存在二次连续可微函数 $v(\cdot)$: $R^l\to R$，使对任意 $\Delta>\delta>0$ 有

$$\sup_{\Delta>d(x,J)>\delta} h^\tau(x)v_x(x)<0.\tag{5.2.4}$$

b) $v(J)=$常数；或者对任一 x，只要 $d(x,J)>0$，必有 $d(v(x),v(J))>0$，这里 $v(J)$ 表示集合 $\{y: y=v(x), \forall x\in J\}$.

c) 对固定的 $x^*\in R^l$，存在实数 M，$\|x^*\|<M$，使

$$v(x^*)\leqslant\inf\{v(x):\|x\|=M\},\tag{5.2.5}$$

并且

$$(\bar{v}(J))^c\cap(v(x^*),\ \inf_{\|x\|=M}v(x))\neq\phi.\tag{5.2.6}$$

A 5.2.3 $\mathscr{F}_t\uparrow$，右连续，$(w_t,\ \mathscr{F}_t)$为 l 维 Wiener 过程. η_t 和$\mathscr{F}_t$适应，并且只要当 $k\to\infty$ 时，$t_k\to\infty$，x_{t_k}收敛，就有

$$\lim_{T\to0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^t a_s\eta_s\,\mathrm{d}s\right\|=0,\ \forall t\in[t_k,\ s(t_k,\ T)].\tag{5.2.7}$$

这一组条件和 A 2.4.0～A 2.4.3 相类似. 由于我们只要求 $h(\cdot)$满足局部 Lipschitz 条件，所以对它的增长速度没有限制. 当 J 是一个单点时，即 $J=x^0$，那么 $v(J)$是常数，所以 A5.2.2 中 b) 自然成立. 如果 $v(x)\xrightarrow[\|x\|\to\infty]{}\infty$，那么 A5.2.2 的 c) 中的 M 一定存在. 这时如果 J 不稠于一个线段，那么 A5.2.2 的 c) 一定成立，

现在来定义变界截尾的 RM 算法. 取单调上升的实数列 $\{M_i\}$，$M_1>1$，当 $i\to\infty$ 时，$M_i\to\infty$.

由于 $h(\cdot)$及 $F(\cdot)$满足局部 Lipschitz 条件，所以下列随机微分方程从初始时刻 t_0 起到爆炸时间 σ_0 存在唯一强解 $x_t^{(1)}$（定理 1.3.3）：

$$x_t^{(1)}=x_{t_0}+\int_{t_0}^t a_s(h(x_s^{(1)})+\eta_s)\mathrm{d}s+\int_{t_0}^t a_sF(x_s^{(1)})\,\mathrm{d}w_s,\ t\in[t_0,\ \sigma_0).\tag{5.2.8}$$

定义停时 α_1：

$$\alpha_1=\begin{cases}\inf\{t:t\geqslant t_0,\ \|x_t^{(1)}\|>M_1\},\\ \infty,\text{如果}\|x_t^{(1)}\|\leqslant M_1,\ \forall t\geqslant t_0.\end{cases}$$

下面的随机微分方程到爆炸时刻 σ_1 止也有唯一强解 $x_t^{(2)}$

$$x_t^{(2)}=x^*+\int_{t_0}^t a_s(h(x_s^{(2)})+\eta_s)I_{[\alpha_1<s]}\,\mathrm{d}s$$

$$+\int_{t_0}^t a_sF(x_s^{(2)})I_{[\alpha_1<s]}\,\mathrm{d}w_s$$

$$=x^*+\int_{\alpha_1}^t a_s\,\mathrm{d}y_s,\quad t<\sigma_1.$$

类似地定义停时 α_2:

$$\alpha_2=\begin{cases}\inf\{t:t\geqslant\alpha_1,\ \|x_t^{(2)}\|>M_2\},\\ \infty,\text{如果 }\|x_t^{(2)}\|\leqslant M_2,\ \forall t\geqslant\alpha_1.\end{cases}$$

这样，我们可以递推地定义下去，定义了停时 α_i 后，我们把算法从 α_i 开始，以 x^* 为初值继续下去，也就是在爆炸时间 σ_i 前，定义 $x_t^{(i+1)}$:

$$\begin{aligned}x_t^{(i+1)}&=x^*+\int_{t_0}^t a_s(h(x_s^{(i+1)})+\eta_s)I_{[\alpha_i<s]}\mathrm{d}s\\&\quad+\int_{t_0}^t a_sF(x_s^{(i+1)})I_{[\alpha_i<s]}\,\mathrm{d}w_s\\&=x^*+\int_{\alpha_i}^t a_s\,\mathrm{d}y_s,\quad t<\sigma_i,\end{aligned}\tag{5.2.9}$$

然后定义停时 α_{i+1}:

$$\alpha_{i+1}=\begin{cases}\inf\{t:t\geqslant\alpha_i,\ \|x_t^{(i+1)}\|>M_{i+1}\},\\ \infty,\text{如果 }\|x_t^{(i+1)}\|\leqslant M_{i+1},\ \forall t\geqslant\alpha_i,\end{cases}\tag{5.2.10}$$

$x_t^{(i+1)}$ 的爆炸时间为 $\sigma_{i+1}(>\alpha_{i+1})$,

记

$$x_t=\sum_{i=1}^{\infty}x_t^{(i)}I_{[\alpha_{i-1}<t<\alpha_i]}.\tag{5.2.11}$$

x_t 就是变界截尾的 RM 算法对 $h(\cdot)$ 的零集 J 的估计.

引理 5.2.1 x_t 在 $[0,\infty)$ 上 a.s. 有定义，也就是说 $\alpha_n\xrightarrow[n\to\infty]{}\alpha$, $\alpha=\infty$ a.s..

证明 由于 $M_n\xrightarrow[n\to\infty]{}\infty$，所以当 $n\geqslant n_0$ 时，$M_n>\|x^*\|$.

定义

$$\alpha'_{i+1}=\begin{cases}\inf\{t:t\geqslant\alpha_i,\ \|x_t^{(i+1)}\|>M_{n_0}\},\\ \infty,\text{如果 }\|x_t^{(i+1)}\|\leqslant M_{n_0},\ \forall t\geqslant\alpha_i,\ \forall i\geqslant n_0-1.\end{cases}\tag{5.2.12}$$

从上面定义知 $\alpha_{n_0}=\alpha'_{n_0}$, $\alpha_{i+1}>\alpha'_{i+1}$, $\forall i\geqslant n_0$.

定义

$$F_s=\begin{cases}F(x_s^{(i)}), & \alpha_{i-1}\leqslant s\leqslant\alpha_i' \\ 0, & \alpha_i'<s<\alpha_i\end{cases}\quad \forall i\geqslant n_0.$$

由于 α_i, α_i' 都是停时，所以 F_s 对 $\mathscr{F}_s$ 适应。

记

$$h=\max_{\|x\|\leqslant M_{n_0}}\|h(x)\|,\quad f=\sup_{x\in R^l}\|F(x)\|.$$

由于

$$\int_{\alpha_{n_0}}^{\alpha}a_s^2\|F_s\|^2\mathrm{d}s\leqslant f^2\int_0^{\infty}a_s^2\mathrm{d}s<\infty,$$

所以 $\int_{\alpha_{n_0}}^{\alpha}a_sF_s\mathrm{d}w_s<\infty$ a.s.，也就是说，当 $n\to\infty$ 时，

$$\int_{\alpha_{n_0}}^{\alpha_n}a_sF_s\,\mathrm{d}w_s=\sum_{i=n_0}^{n}\int_{\alpha_{i-1}}^{\alpha_i'}a_sF(x_s^{(i)})\,\mathrm{d}w_s$$

收敛到有穷极限 a.s.。

对 $\omega\in[\alpha<\infty]$，我们有 $\alpha_i<\infty$，$\alpha_i'<\infty$，$\forall i$，并且 $\alpha_i'-\alpha_{i-1}\xrightarrow[i\to\infty]{}0$. 由式(5.2.9)及式(5.2.12)知

$$\left\|x^*+\int_{\alpha_{i-1}}^{\alpha_i'}a_sh(x_s^{(i)})\,\mathrm{d}s+\int_{\alpha_{i-1}}^{\alpha_i'}a_s\eta_s\,\mathrm{d}s+\int_{\alpha_{i-1}}^{\alpha_i'}a_sF(x_s^{(i)})\mathrm{d}w_s\right\|$$
$$=M_{n_0},$$

由此导致矛盾：

$$0<M_{n_0}-\|x^*\|\leqslant h\int_{\alpha_{i-1}}^{\alpha_i'}a_s\,\mathrm{d}s+\left\|\int_{\alpha_{i-1}}^{\alpha_i'}a_sF_s\,\mathrm{d}w_s\right\|+\int_{\alpha_{i-1}}^{\alpha_i'}a_s\|\eta_s\|\mathrm{d}s$$
$$\xrightarrow[i\to\infty]{}0.$$

这说明 $P[\alpha<\infty]=0$. 所以 $\alpha=\infty$ a.s.。 □

引理 5.2.2 设条件 A5.2.0、A5.2.1 和 A5.2.3 成立，对固定的样本，设 $\{x_{t_n}, n=1, 2, \cdots\}$ 是 $\{x_t\}$ 的收敛子序列. 存在常数 $c_1>0$, $c_2>0$, $\Delta>0$ 及 N_T, 使对任一 $T\in[0, \Delta)$ 及任一 $s\in[t_n, s(t_n, T)]$，只要 $n\geqslant N_T$ 就有

$$\left\|\int_{t_n}^{s}a_\lambda h(x_\lambda)\mathrm{d}\lambda+\int_{t_n}^{s}a_\lambda\,\mathrm{d}e_\lambda\right\|\leqslant c_1,\qquad(5.2.13)$$

$$\|x_s-x_{t_n}\|\leqslant c_2T.\qquad(5.2.14)$$

证明 在下面的讨论中，都针对引理条件中的固定样本. 如

果截尾次数有穷，即存在 i，使 $\alpha_i=\infty$，那么据定义 $\|x_t\|\leqslant M_i$，$\forall t\geqslant t_0$. 所以

$$x_t=x_{\alpha_{i-1}}+\int_{\alpha_{i-1}}^{t}a_s h(x_s)\,\mathrm{d}s+\int_{\alpha_{i-1}}^{t}a_s\,\mathrm{d}\varepsilon_s. \tag{5.2.15}$$

利用 $\|x_t\|$ 的有界性，式(5.2.13)和(5.2.14)显然成立.

现设 $\alpha_i<\infty$，$\forall i$. 记 $\bar{x}=\lim\limits_{n\to\infty}x_{t_n}$.

反设式(5.2.13)不成立. 取 $c>\|\bar{x}\|$. 对 $T=2^{-m}$，不论 m 取何正整数，必存在 k_m 和 s_m，使

$$\left\|\int_{t_{km}}^{s_m}a_\lambda h(x_\lambda)\mathrm{d}\lambda+\int_{t_{km}}^{s_m}a_\lambda\,\mathrm{d}\varepsilon_\lambda\right\|>\frac{c-\|\bar{x}\|}{2}, \tag{5.2.16}$$

$$\|x_{t_n}\|\leqslant\frac{c+\|\bar{x}\|}{2},\quad \forall n\geqslant k_m. \tag{5.2.17}$$

不失一般性可认为

$$s_m=\inf\left\{t;\,t>t_{k_m},\ \left\|\int_{t_{km}}^{t}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_{km}}^{t}a_\lambda\,\mathrm{d}\varepsilon_\lambda\right\|\geqslant\frac{c-\|\bar{x}\|}{2}\right\}, \tag{5.2.18}$$

那么对任意 $s\in[t_{k_m},\ s_m]$，只要 m 充分大，从式(5.2.16)和(5.2.17)知

$$\left\|x_{t_{km}}+\int_{t_{km}}^{t}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_{km}}^{s}a_\lambda\,\mathrm{d}\varepsilon_\lambda\right\|\leqslant c. \tag{5.2.19}$$

由于 $M_i\to\infty$，所以当 i 充分大时，$M_i>c$，$\forall i\geqslant i_0$. 所以从式(5.2.19)及(5.2.9)看出 $\alpha_i>s_m$，也就是说，只要 m 充分大，在 $[t_{k_m},\ s_m]$ 区间内不发生截尾.

$$x_s=x_{t_{k_m}}+\int_{t_{k_m}}^{s}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_{k_m}}^{s}a_\lambda\,\mathrm{d}\varepsilon_\lambda,\ s\in[t_{\lambda_m},\ s_m]. \tag{5.2.20}$$

从式(5.2.19)和(5.2.20)看出 $\|x_s\|$ 在 $[t_{k_m},\ s_m]$ 区间上有界，所以 $h(x_s)$ 也在该区间上有界.

注意到 $s_m\leqslant s(t_{k_m},\ 2^{-m})$，$\displaystyle\int_{t_{\lambda m}}^{s_m}a_\lambda\,\mathrm{d}s\leqslant 2^{-m}\xrightarrow[m\to\infty]{}0$，所以

$$\|x_{s_m}-x_{t_{km}}\|\xrightarrow[m\to\infty]{}0, \tag{5.2.21}$$

另一方面，从式(5.2.18)及(5.2.20)知

$$\|x_{t_m}-x_{t_{km}}\|=\frac{c-\|\bar{x}\|}{2}.$$

这和式(5.2.21)矛盾．这就证明了式(5.2.13)．

设 j 已充分大，使对式(5.2.13)的 c_1 有

$$2c_1+\|\bar{x}\|<M_j.$$

由于对充分大的 n,

$$\|x_{t_n}\|\leqslant\|\bar{x}\|+c_1, \tag{5.2.22}$$

那么由式(5.2.13)知对任意 $s\in[t_n,\ s(t_n,\ T)]$ 有

$$\left\|x_{t_n}+\int_{t_n}^{s}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_n}^{s}a_\lambda\,\mathrm{d}e_\lambda\right\|<M_j. \tag{5.2.23}$$

因此 $\sigma_j\geqslant s(t_n,\ T)$，从而对充分大的 n 有

$$x_s=x_{t_n}+\int_{t_n}^{s}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_n}^{s}a_\lambda\,\mathrm{d}e_\lambda,\ \forall s\in[t_n,\ s(t_n,\ T)]. \tag{5.2.24}$$

从式(5.2.22)及(5.2.23)便知在区间$[t_n,\ s(t_n,\ T)]$上 x_s 及 $h(x_s)$ 都有界，用 k_1 表示它们的公共上界，k_1 不依赖于 n.

由于式(5.2.7)，只要 n 充分大,

$$\left\|\int_{t_n}^{s}a_\lambda\eta_\lambda\,\mathrm{d}\lambda\right\|\leqslant k_1T,\ \forall s\in[t_n,\ s(t_n,\ T)].$$

注意到 $\int_{t_0}^{\infty}a_\lambda F(x_\lambda)\,\mathrm{d}w_\lambda<\infty$ a. s.，所以只要 n 充分大，就有 $\left\|\int_{t_n}^{s}a_\lambda F(x_\lambda)\,\mathrm{d}w_\lambda\right\|<T,\ \forall s\in[t_n,\ s(t_n,\ T)]$．那么从式(5.2.24)就有

$$\|x_s-x_{t_n}\|\leqslant 2k_1\int_{t_n}^{s}a_\lambda\,\mathrm{d}\lambda+T\leqslant(2k_1+1)T.$$

这就证明了式(5.2.14)． □

引理 5.2.3 设条件 A5.2.0～A5.2.3 成立．又设$[\delta_1,\ \delta_2]$，$\delta_1\leqslant\delta_2$，是一个区间，并且 $d([\delta_1,\ \delta_2],\ v(J))>0$．那么 i)当 $\delta_1<\delta_2$ 时，$v(x_t)$不可能无穷次穿越$[\delta_1,\ \delta_2]$(即不可能有无穷多个 k，使 $v(x_{t_k})=\delta_1,\ v(x_{s_k})=\delta_2,\ t_k<s_k,\ \delta_1<v(x_t)<\delta_2,\ \forall t\in(t_k,\ s_k)$)，同时又存在常数 $c>0$，使$\|x_t\|\leqslant c,\ \forall t\in[t_k,\ s_k],\ \forall k$;

ii) 当 $\delta_1=\delta_2$ 时，$v(x_t)$不可能收敛到 δ_1，同时 $\{x_t\}$ 有界。

证明 反设 i)不成立. 也就是设存在$[\delta_1, \delta_2]$, $\delta_2>\delta_1$, $d([\delta_1, \delta_2], v(J))>0$, $v(x_t)$无穷次穿越$[\delta_1, \delta_2]$，并且$\|x_t\|\leqslant c$, $\forall t\in[t_k, s_k]$, $\forall k$, $s_k>t_k \xrightarrow[k\to\infty]{} \infty$.

由于 $\|x_{t_k}\|$ 有界，选它的收敛子序，仍用 x_{t_k} 表示：$x_{t_k}\xrightarrow[k\to\infty]{}\bar{x}$.

用 Taylor 展开得

$$\begin{aligned}v(x_{s(t_k,T)})-v(x_{t_k})&=(x_{s(t_k,T)}-x_{t_k})^\tau v_x(\zeta)\\&=(x_{s(t_k,T)}-x_{t_k})^\tau v_x(\bar{x})\\&\quad+(x_{s(t_k,T)}-x_{t_k})^\tau v_{xx}(\xi)(\zeta-\bar{x}),\end{aligned}\tag{5.2.25}$$

其中 $\|\zeta-x_{t_k}\|\leqslant c_2T$. 当 k 充分大时 $\|x_{t_k}-\bar{x}\|<c_2T$, 所以 $\|\zeta-\bar{x}\|<2c_2T$，于是 $\|\xi-\bar{x}\|<2c_2T$. 所以存在常数 c_3 使 $\|v_{xx}(\xi)\|\leqslant c_3$. 常数 c_2, c_3 都不依赖于 k.

从引理 5.2.2 知，对充分大的 n，对 $s\in[t_n, s(t_n, T)]$，有

$$x_s=x_{t_n}+\int_{t_n}^{s}a_\lambda h(x_\lambda)\,\mathrm{d}\lambda+\int_{t_n}^{s}a_\lambda\,\mathrm{d}e_\lambda.\tag{5.2.26}$$

把式(5.2.26)代入式(5.2.25)并用式(5.2.14)就有

$$\begin{aligned}&v(x_{s(t_k,T)})-v(x_{t_k})\\&\leqslant\int_{t_k}^{s(t_k,T)}a_\lambda h^\tau(\bar{x})v_x(\bar{x})\mathrm{d}\lambda+\int_{t_k}^{s(t_k,T)}a_\lambda(h(x_\lambda)-h(\bar{x}))^\tau v_x(\bar{x})\mathrm{d}\lambda\\&\quad+\int_{t_k}^{s(t_k,T)}a_\lambda\,\mathrm{d}\varepsilon_\lambda^\tau v_x(\bar{x})+c_4T^2,\ c_4\text{ 为常数.}\end{aligned}\tag{5.2.27}$$

因为 $x_{t_k}\to\bar{x}$，所以 $v(\bar{x})=\delta_1$. 据引理条件$d(v(\bar{x}), v(J))>0$，所以$d(\bar{x}, J)>0$. 据 A5.2.2 中 a)

$$h^\tau(\bar{x})v_x(\bar{x})=-a<0.$$

那么从式(5.2.27)用 $\|F(\cdot)\|$ 的有界性及 A5.2.3 就得

$$\begin{aligned}&v(x_{s(t_k,T)})-v(x_{t_k})\\&\leqslant-aT+a_4T^2+\int_{t_k}^{s(t_k,T)}a_\lambda(h(x_\lambda)-h(\bar{x}))^\tau v_x(\bar{x})\mathrm{d}\lambda\\&\quad+o(T),\end{aligned}\tag{5.2.28}$$

这里 $o(T)\xrightarrow[\substack{k\to\infty\\T\to 0}]{}0$.

由式(5.2.14)知

$$\left\|\int_{t_k}^{s(t_k,T)} a_\lambda (h(x_\lambda)-h(\bar{x}))^\tau v_x(\bar{x})\mathrm{d}\lambda\right\|\leqslant o(T),\ \forall k.$$

在式(5.2.28)中把 $k\to\infty$, 便知存在 $a'>0$, 使

$$\limsup_{k\to\infty} v(x_{s(t_k,T)})\leqslant \delta_1-a'T. \tag{5.2.29}$$

另一方面, 从式(5.2.14)知

$$\max_{t_k\leqslant t\leqslant s(t_k,T)} |v(x_t)-v(x_{t_k})|\xrightarrow[T\to 0]{}0. \tag{5.2.30}$$

由于 $v(x_{t_k})=\delta_1$, $v(x_{s_k})=\delta_2$, $\delta_1<v(x_t)<\delta_2, t\in(t_k, s_k)$, 从式(5.2.30)便知对充分小的 T, $s(t_k, T)\leqslant s_k$. 也就是说, $v(x_{s(t_k,T)})\in[\delta_1, \delta_2]$. 这和式(5.2.29)矛盾. 所以 i)成立.

ii) 当 $\{x_t\}$ 有界时, 则有收敛子列 $x_{t_k}\xrightarrow[k\to\infty]{}\bar{x}$. 则式(5.2.25)～(5.2.29)仍成立, 但据 ii)的反假设, 式(5.2.29)的左端收敛到 δ_1, 从而得出矛盾. □

定理5.2.1 设 A5.2.0～A5.2.3 成立. 那么由式(5.2.11)定义的 x_t a.s. 收敛到 J, 即

$$\lim_{t\to\infty}\mathrm{d}(x_t, J)=0.$$

证明 据式(5.2.5)、(5.2.6)存在 $\delta_1\in(v(x^*), \inf_{\|x\|=M} v(x))$, 使 $d(\delta_1, v(J))>0$. 所以存在 $\delta_2>\delta_1$, 使 $d([\delta_1, \delta_2], v(J))>0$.

我们来证, 对固定的样本必存在 i 使 $a_i=\infty$. 否则 x_t 将无穷次回到 x^*, 并穿越球 $\{x: \|x\|=M\}$. 由于 $\|x^*\|<M$, 所以 $v(x_t)$ 将无穷次穿越区间$[\delta_1, \delta_2]$, 即存在 $t_k<s_k$, $t_k\to\infty$, 使 $v(x_{t_k})=\delta_1$, $v(x_{s_k})=\delta_2$. 同时对 $t\in[t_k, s_k]$还有 $\|x_t\|\leqslant M$. 据引理5.2.3的 i), 这不可能发生, 所以对 $t\geqslant a_{i-1}$, 算法和 RM 重合, $\|x_t\|\leqslant M_i$.

我们来证

$$\lim_{t\to\infty} d(v(x_t), v(J))=0 \quad \text{a.s.} \tag{5.2.31}$$

记

$$v_1=\liminf_{t\to\infty} v(x_t)\leqslant\limsup_{t\to\infty} v(x_t)=v_2. \qquad (5.2.32)$$

如果 $v_1=v_2$，据引理 5.2.2 的 ii)，$d(v_1, v(J))=0$. 所以式(5.2.31)成立.

如果 $v_1<v_2$，并且至少有一个 v_i, $i=1, 2$ 不属于 $\bar{v}(J)$. 不失一般性设 $v_1\overline{\in}\,\bar{v}(J)$. 那么必存在充分小的 $\varepsilon>0$，使

$$[v_1+\varepsilon,\ v_1+2\varepsilon]\cap\bar{v}(J)=\varnothing,$$

且 $v_1+2\varepsilon<v_2$. 从式(5.2.32)知 $v(x_t)$ 无穷次穿越区间 $[v_1+\varepsilon, v_1+2\varepsilon]$. 根据引理 5.2.2 的 i)这不可能，所以 $v_1\in\bar{v}(J)$，$v_2\in\bar{v}(J)$. 同理可证 $v(x_t)$的任一极限点 $\in\bar{v}(J)$，所以式(5.2.31)成立.

现在利用条件 A 5.2.2 的 b)来证 $d(x_t, J)\xrightarrow[t\to\infty]{}0$ a.s..

如果 $v(J)=$常数，记作 δ_1，据式(5.2.31)有

$$v(x_t)\xrightarrow[t\to\infty]{}\delta_1.$$

反设存在子列 $x_{t_k}\to\bar{x}$, $d(\bar{x}, J)>0$. 那么式(5.2.25)～(5.2.29)仍成立. 而它是一个矛盾的不等式:

$$\delta_1\leqslant\delta_1-\alpha' T,$$

所以 $d(x_t, J)\xrightarrow[t\to\infty]{}0$ a.s..

如果对任一 x，只要 $d(x, J)>0$，就有 $d(v(x), v(J))>0$. 设 x_{t_k} 为 x_t 的任一收敛子列，$x_{t_k}\xrightarrow[k\to\infty]{}\bar{x}$. 从式(5.2.31)知 $d(\bar{x}, v(J))=0$，所以 $d(\bar{x}, J)=0$，从 $\{x_{t_k}\}$ 的任意性，便知 $d(x_t, J)\xrightarrow[t\to\infty]{}0$ a.s..

□

注 5.2.2 和定理 3.1.2 类似，如果 $f(\cdot)$ 是 $R^l\to R$ 的二次连续可微函数，J 为 $f(\cdot)$ 的极值点，$f(x)=\text{const}$, $x\in J$. 在定理 5.2.1 中 $h(\cdot)=f_x(\cdot)$，那么满足式(5.2.4)的 $v(\cdot)$ 可取为 $-f(\cdot)$.

§5.3 收敛速度

在上一节中我们证明了由式(5.2.11)定义的 x_t 收敛到 $h(\cdot)$

的零点集 J. 这一节中我们来考察 J 为一个单点 x^0 时的收敛速度.

我们要用以下条件.

A 5.3.0 $\|F(\cdot)\|$ 有界, $f=\sup_{x\in R^l}\|F(x)\|$, $h(\cdot)$ 及 $F(\cdot)$ 满足局部 Lipschitz 条件, $h(x^0)=0$, 并且存在稳定阵 H 使

$$\begin{aligned}&h(x)=H(x-x^0)+\delta(x),\ \delta(x^0)=0,\\&\text{当 } x\to x^0 \text{ 时},\quad \delta(x)=o(\|x-x^0\|).\end{aligned}\tag{5.3.1}$$

A 5.3.1 确定性的增益函数 $\{a_t\}$ 满足以下条件: a_t 非增,

$$a_t>0,\ a_t\xrightarrow[t\to\infty]{}0,\ \int_{t_0}^{\infty}a_t\,\mathrm{d}t=\infty,\ \int_{t_0}^{\infty}a_t^{2(1-\delta)}\,\mathrm{d}t<\infty,\ \delta\geqslant 0,$$

并且

$$\frac{1}{(t-s)}(a_t^{-1}-a_s^{-1})\xrightarrow[s\to\infty]{}\alpha\geqslant 0,\ t>s.\tag{5.3.2}$$

$\left(\text{如果 } a_t=\dfrac{1}{t}, \text{则 } \alpha=1,\ \delta\in\left[0,\ \dfrac{1}{2}\right)\right)$.

A 5.3.2 存在二次连续可微 $R^l\to R$ 函数 $v(\cdot)$, 使对任意 $\Delta_2>\Delta_1>0$ 有

$$\sup_{\Delta_1<\|x-x^0\|<\Delta_2}h^{\tau}(x)v_x(x)<0.\tag{5.3.3}$$

还设存在常数 $M>0$, $\|x^*\|<M$, 使 $v(x^*)\neq\inf_{\|x\|=M}v(x)$.

A 5.3.3 $\mathscr{F}_t\uparrow$ 右连续, $(w_t,\mathscr{F}_t)$ 为 l 维 Wiener 过程 η_t 和 $\mathscr{F}_t$ 适应, $\eta_t=o(a_t^{\delta})$ a.s..

定理 5.3.1 设 A5.3.0~A5.3.3 成立, 并且 $H+\alpha\delta I$ 是稳定阵, 那么由式(5.2.11)给出的 x_t 以如下速度收敛到 x^0:

$$\|x_t-x^0\|=o(a_t^{\delta})\quad \text{a.s.}.\tag{5.3.4}$$

证明 由式(5.3.2)知

$$\frac{a_s-a_t}{a_sa_t(t-s)}\xrightarrow[s\to\infty]{}\alpha,$$

所以

$$\mathrm{d}a_t=-\alpha a_t^2\,\mathrm{d}t+o(1)a_t^2\,\mathrm{d}t.\tag{5.3.5}$$

由于定理 5.2.1 的条件满足, 所以 $x_t\to x^0$, 并且存在 t_0, 对 $t>$

a_{i_0}(i_0 依赖于样本),

$$dx_t = a_t h(x_t)dt + a_t d\varepsilon_t. \tag{5.3.6}$$

不失一般性可认为 $x^0=0$, 即 $x_t \xrightarrow[t\to\infty]{} 0$ a.s., 和式(3.2.22)类似地可把 $\delta(x_t)$ 写成 $\delta(x_t)=D_t x_t$, $D_t=\delta(x_t)\cdot\frac{x_t^T}{\|x_t\|^2}\xrightarrow[t\to\infty]{} 0$. 记

$$z_t = a_t^{-\delta} x_t. \tag{5.3.7}$$

从式(5.3.5)和(5.3.6)得

$$\begin{aligned} dz_t &= \alpha\delta a_t z_t\,dt + o(1)a_t z_t\,dt + a_t^{1-\delta}h(x_t)dt + a_t^{1-\delta}d\varepsilon_t \\ &= a_t H_t z_t\,dt + a_t^{1-\delta}d\varepsilon_t, \end{aligned} \tag{5.3.8}$$

这里

$$H_t = H + \alpha\delta I + o(1) \xrightarrow[t\to\infty]{} H + \alpha\delta I. \tag{5.3.9}$$

用 $\Phi_{t,s}$, $\Psi_{t,s}$ 表示基本解阵,

$$\frac{d}{dt}\Phi_{t,s} = a_t H_t \Phi_{t,s},\ \Phi_{s,s}=I,\quad t\geqslant s\geqslant t_0, \tag{5.3.10}$$

$$\frac{d}{dt}\Psi_{t,s} = a_t(H+\alpha\delta I)\Psi_{t,s},\ \Psi_{s,s}=I. \tag{5.3.11}$$

容易验证

$$\Psi_{t,s} = \exp\left[(H+\alpha\delta I)\int_s^t a_\lambda d\lambda\right], \tag{5.3.12}$$

由于 $H+\alpha\delta I$ 稳定, 所以

$$\|\Psi_{t,s}\| \leqslant c_0 \exp\left(-c\int_s^t a_\lambda d\lambda\right),\ c_0>0,\ c>0,\quad \forall t\geqslant s\geqslant 0, \tag{5.3.13}$$

记

$$f_{t,s} = \exp\left(c\int_s^t a_\lambda d\lambda\right)\|\Phi_{t,s}\|. \tag{5.3.14}$$

从式(5.3.10)~(5.3.13)易知

$$\frac{d}{dt}\Phi_{t,s} = a_t(H+\alpha\delta I)\Phi_{t,s} + a_t(H_t - H - \alpha\delta I)\Phi_{t,s},$$

$$\Phi_{t,s} = \Psi_{t,s} + \int_s^t \Psi_{t\lambda} a_\lambda (H_\lambda - H - \alpha\delta I)\Phi_{\lambda,s}\,d\lambda, \tag{5.3.15}$$

$$\|\Phi_{t,s}\| \leqslant c_0 \exp\left(-c\int_s^t a_\lambda d\lambda\right)$$
$$+\int_s^t c_0\left[\exp\left(-c\int_\lambda^t a_\mu d\mu\right)\right] a_\lambda o(1)\|\Phi_{\lambda,s}\| d\lambda.$$

所以
$$f_{t,s} \leqslant c_0 + \int_s^t o(1) a_\lambda f_{\lambda,s} d\lambda.$$

用 Bellman-Grownwall 不等式知

$$f_{t,s} \leqslant c_0 \exp\left(\int_s^t o(1) a_\lambda d\lambda\right).$$

因此
$$\|\Phi_{t,s}\| \leqslant c_0 \exp\left(\int_s^t (o(1)-c) a_\lambda d\lambda\right),$$

这里 $o(1) \xrightarrow[\lambda\to\infty]{} 0$. 所以存在常数 $c_1>0$, $c_2>0$. 使

$$\|\Phi_{t,s}\| \leqslant c_1 \exp\left(-c_2\int_s^t a_\lambda d\lambda\right). \tag{5.3.16}$$

从式(5.3.8)知对 $t \geqslant t_1 > a_{t_0}$

$$z_t = \Phi_{t,t_1} z_{t_1} + \int_{t_1}^t \Phi_{t,s} a_s^{1-\delta} d\varepsilon_s$$
$$= \Phi_{t,t_1} z_{t_1} + \int_{t_1}^t \Phi_{t,s} a_s^{1-\delta}(F(x_s) dw_s + \eta_s ds),$$
$$\|z_t\| \leqslant c_1 \exp\left(-c_2\int_{t_1}^t a_\lambda d\lambda\right)\|z_{t_1}\|$$
$$+c_1\int_{t_1}^t\left[\exp\left(-c_2\int_s^t a_\lambda d\lambda\right)\right] a_s o(1) ds$$
$$+\int_{t_1}^t \Phi_{t,s} a_s^{1-\delta} F(x_s) dw_s. \tag{5.3.17}$$

上式右端第一项，当 $t\to\infty$ 时趋于 0. 记

$$g_{t,s} = \int_s^t a_\lambda d\lambda,$$

那么式(5.3.17)右端第二项为

$$c_1\left[\exp\left(-c_2\int_{t_1}^t a_\lambda d\lambda\right)\right]\int_{t_1}^t\left[\exp\left(c_2\int_{t_1}^s a_\lambda d\lambda\right)\right] a_s o(1) ds$$
$$= c_1\left[\exp(-c_2 g_{t,t_1})\right]\int_{t_1}^t\left[\exp(c_2 g_{s,t_1})\right] o(1) dg_{s,t_1}$$

$$=\frac{c_1}{c_2}\left[\exp(-c_2 g_{t,t_1})\right]\int_{t_1}^{t} o(1)\,\mathrm{d}\exp(c_2 g_s,\ t_1)\xrightarrow[t\to\infty]{} 0. \tag{5.3.18}$$

由于

$$\int_{t_1}^{\infty} a_s^{2(1-\delta)}\|F(x_s)\|^2\mathrm{d}s\leqslant f^2\int_{t_1}^{\infty} a_s^{2(1-\delta)}\mathrm{d}s<\infty,$$

所以

$$G_t=\int_{t_1}^{t} a_s^{1-\delta}F(x_s)\,\mathrm{d}w_s\xrightarrow[t\to\infty]{} G\triangleq\int_{t_1}^{\infty} a_s^{1-\delta}F(x_s)\,\mathrm{d}w_s$$
$$<\infty\quad \text{a.s.}. \tag{5.3.19}$$

因此对式(5.3.17)的最后一项估计如下

$$\Phi_{t,0}\int_{t_1}^{t}\Phi_{0,s}a_s^{1-\delta}F(x_s)\,\mathrm{d}w_s$$
$$=\Phi_{t,0}\int_{t_1}^{t}\Phi_{0,s}\,\mathrm{d}G_s=G_t-\Phi_{t,0}\int_{t_1}^{t}\Phi_{0,s}a_sH_sG_s\,\mathrm{d}s$$
$$=G_t-G+\Phi_{t,t_1}G-\Phi_{t0}\int_{t_1}^{t}\Phi_{0,s}a_sH_s(G_s-G)\,\mathrm{d}s. \tag{5.3.20}$$

由式(5.3.16)和(5.3.19)知上式右端除最后一项外，当 $t\to\infty$ 时趋于 0. 由式(5.3.9)知 H_s 有界，注意到式(5.3.16)便知存在常数 $c_3>0$ 使

$$\left\|\int_{t_1}^{t}\Phi_{t,s}a_sH_s(G_s-G)\,\mathrm{d}s\right\|$$
$$\leqslant c_3\int_{t_1}^{t}\exp\left(-c\int_{s}^{t}a_\lambda\mathrm{d}\lambda\right)a_s\|G_s-G\|\,\mathrm{d}s$$
$$=\left[\exp(-cg_{t,t_1})\right]c_3\int_{t_1}^{t}\exp(cg_{s,t_1})\|G_s-G\|\,\mathrm{d}g_{s,t_1}$$
$$=\frac{c_3}{c}\left[\exp(-cg_{t,t_1})\right]\int_{t_1}^{t}\|G_s-G\|\,\mathrm{d}\exp(cg_{s,t_1}).$$

由式(5.3.19)及 $\int_{t_1}^{\infty}a_\lambda\,\mathrm{d}\lambda=\infty$，便知上式右端当 $t\to\infty$ 时趋于 0，所以式(5.3.20)右端当 $t\to\infty$ 时趋于 0. 这说明式(5.3.17)右端诸项当 $t\to\infty$ 时都趋于 0，从 Z_t 的定义(5.3.7)，便知式(5.3.4)成立. □

§5.4 渐近正态性

这一节仍讨论 $h(\cdot)$ 有单个零点 x^0 的情形．我们来讨论由式(5.2.11)给出的 x_t 的渐近正态性，确切地说，我们将证明 $a_t^{-\frac{1}{2}}(x_t-x^0)$ 当 $t\to\infty$ 时，趋于一个正态分布．

我们要用以下条件：

A 5.4.0 $\|F(\cdot)\|$ 有界，$f=\sup\limits_{x\in R^l}\|F(x)\|$，$h(\cdot)$ 及 $F(\cdot)$ 满足局部 Lipschitz 条件，$h(x^0)=0$，当 $x\to x^0$ 时

$$\|h(x)-H(x-x^0)\|\leqslant k\|x+x^0\|^{1+\beta}, \tag{5.4.1}$$

$k>0$ 为常数，$\beta\in(0,\ 1]$ 也是常数．$H+\frac{1}{2}\alpha I$ 为稳定阵，α 的定义见 A5.4.1．

A 5.4.1 a_t 满足以下条件，$a_t>0$，$a_t\xrightarrow[t\to\infty]{}0$，$\int_{t_1}^{\infty}a_t\,\mathrm{d}t=\infty$，$a_t$ 单调，

$$\frac{1}{(t-s)}\ (a_t^{-1}-a_s^{-1})\xrightarrow[s\to\infty]{}\alpha\geqslant 0,\ \ t\geqslant s, \tag{5.4.2}$$

并对某个 η：

$$0<\eta<\frac{\beta}{1+\beta},\qquad \int_{t_1}^{\infty}a_t^{1+\eta}\,\mathrm{d}t<\infty.$$

A 5.4.2 存在二次连续可微 $R^l\to R$ 函数 $v(\cdot)$，使对任意 $\Delta_2>\Delta_1>0$ 有

$$\sup_{\Delta_1<\|x-x^0\|<\Delta_2}h^{\tau}(x)v_x(x)<0, \tag{5.4.3}$$

还设存在常数 $M>0$，$\|x^*\|<M$，使 $v(x^*)<\inf\limits_{\|x\|=M}v(x)$．

A 5.4.3 $\mathscr{F}_t\uparrow$，右连续，$(w_t,\ \mathscr{F}_t)$ 为 l 维 Wiener 过程，η_t 和 $\mathscr{F}_t$ 适应，$\eta_t=o(\sqrt{a_t})$ a.s.．

定理5.4.1 设条件 A 5.4.0～A 5.4.3 成立，x_t 由式(5.2.11)给出，那么

$$\frac{1}{\sqrt{a_t}}\ (x_t-x^0)\xrightarrow[t\to\infty]{d}N(0,\ S), \tag{5.4.4}$$

$$S=\int_0^\infty\left\{\exp\left[\left(H+\frac{1}{2}\alpha I\right)t\right]\right\}F(x^0)F^\tau(x^0)$$

$$\cdot\exp\left[\left(H^\tau+\frac{1}{2}\alpha I\right)t\right]\mathrm{d}t. \tag{5.4.5}$$

证明 由于 $1-\eta>\frac{1}{1+\beta}$，所以可选 δ，使

$$1-\eta>2\delta>\frac{1}{1+\beta}, \tag{5.4.6}$$

所以 $2-2\delta>1+\eta$，条件 A5.3.1 成立，因此定理 5.3.1 所要求的条件全部成立，从而

$$x_t-x^0=o(a_t^\delta).$$

利用式(5.4.1)便知当 $t\to\infty$ 时，

$$\|h(x_t)-H(x_t-x^0)\|\leqslant o(a_t^{\delta(1+\beta)}),\ \delta(1+\beta)>\frac{1}{2}. \tag{5.4.7}$$

记

$$\xi_t=\frac{1}{\sqrt{a_t}}(x_t-x^0). \tag{5.4.8}$$

从式(5.3.5)、(5.3.6)和(5.4.7)并注意到 $\eta_t=o(\sqrt{a_t})$ 可得：对 $t\geqslant t_1>\alpha_{t_0}$，

$$\begin{aligned}\mathrm{d}\xi_t=&-\frac{1}{2}a_t^{-\frac{1}{2}-1}(-\alpha a_t^2\mathrm{d}t+o(1)a_t^2\mathrm{d}t)(x_t-x^0)\\&+a_t^{-\frac{1}{2}}(a_th(x_t)\mathrm{d}t+a_t\mathrm{d}e_t)\\=&\frac{\alpha}{2}a_t\xi_t\mathrm{d}t+o(1)a_t\xi_t\mathrm{d}t+a_tH\xi_t\mathrm{d}t\\&+o(a_t^{\delta(1+\beta)-\frac{1}{2}})a_t\mathrm{d}t+o(1)a_t\mathrm{d}t+\sqrt{a_t}F(x_t)\mathrm{d}w_t\\=&\left(H+\frac{\alpha}{2}I+o(1)\right)a_t\xi_t\mathrm{d}t+o(1)a_t\mathrm{d}t\\&+\sqrt{a_t}F(x^0)\mathrm{d}w_t+\sqrt{a_t}(F(x_t)-F(x^0))\mathrm{d}w_t\\=&H_ta_t\xi_t\mathrm{d}t+o(1)a_t\mathrm{d}t+\sqrt{a_t}F(x^0)\mathrm{d}w_t\\&+\sqrt{a_t}(F(x_t)-F(x^0))\mathrm{d}w_t,\end{aligned} \tag{5.4.9}$$

这里 H_t 是确定性阵：

$$H_t=\left(H+\frac{\alpha}{2}I+o(1)\right)\xrightarrow[t\to\infty]{}H+\frac{\alpha}{2}I. \qquad (5.4.10)$$

在式(5.3.9)中取 $\delta=\frac{1}{2}$，由于已假定 $H+\frac{\alpha}{2}I$ 稳定，所以式(5.3.10)～(5.3.16)仍成立。

从式(5.4.9)知对 $t\geqslant t_1>\alpha_{t_0}$，

$$\xi_t=\Phi_{tt_1}\xi_{t_1}+\int_{t_1}^t\Phi_{t,s}o(1)a_s\,\mathrm{d}s+\int_{t_1}^t\Phi_{t,s}\sqrt{a_s}\,F(x^0)\mathrm{d}w_s$$

$$+\int_{t_1}^t\Phi_{t,s}\sqrt{a_s}\,(F(x_s)-F(x^0))\mathrm{d}w_s. \qquad (5.4.11)$$

由式(5.3.16)知上式右端第一项当 $t\to\infty$ 时，趋于 0，而第二项小于等于

$$c_1\int_{t_1}^t\exp\left[-c_2\int_s^t a_\lambda\mathrm{d}\lambda\right]o(1)a_s\,\mathrm{d}s,$$

我们在式(5.3.18)中已证这个量当 $t\to\infty$ 时趋于0. 对式(5.4.11)右端最后一项取二阶矩，并注意到 $\Phi_{t,s}$ 是确定性阵，用控制收敛定理又知 $\mathrm{E}\|F(x_s)-F(x^0)\|^2=o(1)$，我们就得

$$\mathrm{E}\left\|\int_{t_1}^t\Phi_{t,s}\sqrt{a_s}\,(F(x_s)-F(x^0))\mathrm{d}w^s\right\|^2$$

$$=\mathrm{tr}\int_{t_0}^t\Phi_{t,s}a_s\mathrm{E}I_{[\alpha_{t_0}<s]}(F(x_s)-F(x^0))(F(x_s)$$

$$-F(x^0))^{\tau}\Phi_{ts}^{\tau}\,\mathrm{d}s,$$

$$l\int_{t_0}^t\|\Phi_{t,s}a_so(1)\Phi_{t,s}^{\tau}\|\mathrm{d}s$$

$$\leqslant l\,c_1^2\int_{t_0}^t\left[\exp\left(-2c_2\int_s^t a_\lambda\,\mathrm{d}\lambda\right)\right]a_so(1)\mathrm{d}s\xrightarrow[t\to\infty]{}0,$$

这里 t_0 表示确定性的初始时间，而 $t_1>\alpha_{t_0}$ 依赖于样本. 所以式(5.4.11)的最后一项依概率收敛到0. 那么 ξ_t 和 $\int_{t_1}^t\Phi_{t,s}\sqrt{a_s}\cdot F(x^0)\mathrm{d}w_s$ 的极限分布相同，我们只要证明

$$\int_{t_1}^t\Phi_{t,s}\sqrt{a_s}\,F(x^0)\mathrm{d}w_s\xrightarrow[t\to\infty]{d}N(0,\ S). \qquad (5.4.12)$$

由于$\Phi_{t,s}$是确定性阵，所以上式左端服从正态分布，均值为0，协方差阵为

$$S_t \triangleq \int_{t_1}^{t} \Phi_{t,s} a_s F(x^0) F^{\tau}(x^0) \Phi_{t,s}^{\tau} \mathrm{d}s. \tag{5.4.13}$$

为证定理，我们只要证明 $S_t \xrightarrow[t\to\infty]{} S$.

从式(5.3.13)、(5.3.15)、(5.3.16)知$\left(\delta=\frac{1}{2}\right)$

$$\|\Phi_{t,s}-\Psi_{t,s}\| \leqslant \int_s^t c_0 \left[\exp\left(-c\int_\lambda^t a_\mu \mathrm{d}\mu\right)\right] a_\lambda o(1) c_1 \mathrm{d}\lambda,$$

这里 $o(1)\xrightarrow[\lambda\to\infty]{} 0$. 对任意 $\varepsilon>0$, 可取 s 充分大，使 $\|o(1)\|<\varepsilon$, $\forall \lambda \geqslant s$, 所以从上式和式(5.3.18)类似地有

$$\|\Phi_{t,s}-\Psi_{t,s}\| \leqslant \frac{c_0 c_1}{c}[\exp(-cg_{t,s})]\int_s^t o(1)\mathrm{d}\exp(cg_{\lambda,s})$$

$$\leqslant \frac{\varepsilon c_0 c_1}{c}[\exp(-cg_{t,s})]\int_s^t \mathrm{d}\exp(cg_{\lambda,s}) \leqslant \frac{\varepsilon c_0 c_1}{c}, \forall t\geqslant s.$$

所以对 $t\geqslant s$ 一致地有

$$\|\Phi_{t,s}-\Psi_{t,s}\| \xrightarrow[s\to\infty]{} 0, \tag{5.4.14}$$

这样，我们由式(5.3.13)、(5.3.16)和(5.4.14)就可得：

$$\left\|s_t - \int_{t_1}^t \Psi_{t,s} a_s F(x^0) F^{\tau}(x^0) \Psi_{t,s}^{\tau} \mathrm{d}s\right\|$$

$$\leqslant \left\|\int_{t_1}^t (\Phi_{t,s}-\Psi_{t,s}) a_s F(x^0) F^{\tau}(x^0) \Phi_{t,s} \mathrm{d}s\right\|$$

$$+ \left\|\int_{t_1}^t \Psi_{t,s} a_s F(x^0) F^{\tau}(x^0) (\Phi_{t,s}-\Psi_{t,s}) \mathrm{d}s\right\|$$

$$\leqslant f^2 c_1 \int_{t_1}^t o(1) a_s \exp\left[-c_2\int_s^t a_\lambda \mathrm{d}\lambda\right]\mathrm{d}s + f^2 c_0 \int_{t_1}^t o(1) a_s$$

$$\cdot \exp\left[-c\int_s^t a_\lambda \mathrm{d}\lambda\right]\mathrm{d}s.$$

从式(5.3.18)知上式右端当 $t\to\infty$ 时趋于0. 所以为证 $s_t\to s$, 我们只要证明

$$\int_{t_1}^t \Psi_{t,s} a_s F(x^0) F^{\tau}(x^0) \Psi_{t,s}^{\tau} \mathrm{d}s \to S.$$

仍用 $g_{t,s}$表示 $\int_s^t a_\lambda \mathrm{d}\lambda$, 那么

$$\Psi_{t,s}=\exp\left[\left(H+\frac{\alpha}{2}I\right)g_{t,s}\right].$$

所以,

$$\int_{t_1}^{t}\Psi_{t,s}a_sF(x^0)F^{\tau}(x^0)\Psi_{t,s}^{\tau}\,\mathrm{d}s$$

$$=\int_{t_1}^{t}\exp\left[\left(H+\frac{\alpha}{2}I\right)g_{t,s}\right]\mathrm{d}(g_{t,t_1}-g_{t,s})F(x^0)F^{\tau}(x^0)$$

$$\cdot\exp\left[\left(H+\frac{\alpha}{2}I\right)g_{t,s}\right]$$

$$=-\int_{t_1}^{t}\left\{\exp\left[\left(H+\frac{\alpha}{2}I\right)g_{t,s}\right]\right\}\mathrm{d}g_{t,s}F(x^0)F^{\tau}(x^0)$$

$$\cdot\exp\left[\left(H^{\tau}+\frac{\alpha}{2}I\right)g_{t,s}\right]$$

$$=\int_{0}^{g_{t,t_1}}\left\{\exp\left[\left(H+\frac{\alpha}{2}I\right)s\right]\right\}F(x^0)F^{\tau}(x^0)$$

$$\cdot\exp\left[\left(H^{\tau}+\frac{\alpha}{2}I\right)s\right]\mathrm{d}s$$

由于 $g_{t,t_1}\xrightarrow[t\to\infty]{}\infty$,所以上式当 $t\to\infty$ 时,收敛到 S. □

§5.5 渐近有效性

在式(5.2.8)～(5.2.11)对算法 x_t 的定义中,用 a_tD 来取代 a_t,那么在定理 5.4.1 的条件中,只要把 $H+\frac{\alpha}{2}I$ 稳定改为 $DH+\frac{\alpha}{2}I$ 稳定,而其他条件不变,定理 5.4.1 的结论仍对. 这时

$$\frac{1}{\sqrt{a_t}}(x_t-x^0)\xrightarrow[t\to\infty]{d}N(0,S_D),\qquad(5.5.1)$$

$$S_D=\int_0^{\infty}\left\{\exp\left[\left(DH+\frac{\alpha}{2}I\right)t\right]\right\}DF(x^0)F^{\tau}(x^0)D^{\tau}$$

$$\cdot\exp\left[\left(DH+\frac{\alpha}{2}I\right)^{\tau}t\right]\mathrm{d}t\qquad(5.5.2)$$

根据定理 4.3.1,当 $D=-H^{-1}$ 时,S_D 达到极小,即

$$\min_D S_D=\alpha H^{-1}F(x^0)F^{\tau}(x^0)H^{-\tau}.\qquad(5.5.3)$$

当 $a_t=t^{-1}$ 时，那么从 A5.4.1 中看出，$\alpha=1$，这时取增益系数

$$a_t=\frac{-H^{-1}}{t}, \tag{5.5.4}$$

就有

$$\sqrt{t}\,(x_t-x^0)\xrightarrow[t\to\infty]{d}N(0,\ S),$$

$$S=H^{-1}F(x^0)F^{\tau}(x^0)H^{-\tau}, \tag{5.5.5}$$

如果对 x^0 的估计 $\hat{x}_t$ 使 $\sqrt{t}\,(\hat{x}_t-x^0)\xrightarrow[t\to\infty]{d}N(0,\ S)$，$S$ 由式(5.5.5)给出，那么称 $\hat{x}_t$ 为对 x^0 的渐近有效估计. S 是极限分布的最小协方差阵.

适应随机逼近的基本想法是用对 H 的估计 H_t 来取代 H，取增益系数等于 $\frac{-H_t^{-1}}{t}$. 但这类算法为了保证 H_t 收敛到 H，所用的条件比较苛刻，而计算量也比较大. 下面我们用连续时间的慢衰减增益加平均的方法. 来达到最小的渐近协方差阵(5.5.5). 为此先来定义"慢衰减增益"，并给出它的性质.

如果 $a_t>0$, $t\in(0,\ \infty)$ 当 $t\to\infty$ 时单调非增地趋于 0，并且

$$\frac{1}{t-s}\,(a_t^{-1}-a_s^{-1})\xrightarrow[s\to\infty]{}0,\qquad t>s, \tag{5.5.6}$$

那么称 $\{a_t\}$ 为慢衰减增益.

我们先给出 2 个慢衰减增益的例子.

例 1 $t^{-\alpha}$, $\alpha\in(0,\ 1)$.

记 $\Delta_t=t-s$，我们有

$$\frac{t^\alpha-s^\alpha}{t-s}=\frac{(s+\Delta_t)^\alpha-s^\alpha}{\Delta_t}.$$

当 $\Delta_t\leqslant s$ 时，就有

$$\frac{(s+\Delta_t)^\alpha-s^\alpha}{\Delta_t}\leqslant\frac{s^\alpha}{\Delta_t}\left(\left(1+\frac{\Delta_t}{s}\right)^\alpha-1\right)\leqslant\frac{s^\alpha}{\Delta_t}\left(1+\frac{\alpha\Delta_t}{s}-1\right)$$

$$=\alpha s^{\alpha-1}\xrightarrow[s\to\infty]{}0.$$

当 $\Delta_t>s$ 时，那么

$$\frac{(s+\Delta_t)^\alpha-s^\alpha}{\Delta_t}=\frac{1}{\Delta_t^{1-\alpha}}\left[\left(1+\frac{s}{\Delta_t}\right)^\alpha-\left(\frac{s}{\Delta_t}\right)^\alpha\right]$$

$$\leqslant \frac{1}{s^{1-\alpha}}\left(1+\frac{\alpha s}{\Delta_t}\right)\leqslant(1+\alpha)\,s^{\alpha-1}\xrightarrow[s\to\infty]{}0.$$

例 2 设 $a_t>0$, $t\in(0,\ \infty)$, 并且当 $t\to\infty$ 时单调地收敛到 0. 如果对任意 $\delta\in(0,\ 1)$, 有

$$\limsup_{t\to\infty}\frac{a_t}{a_{(1+\delta)t}}\leqslant 1+q\delta, \tag{5.5.7}$$

q 为不依赖于 δ 的常数, $q\in(0,\ 1)$. 那么式(5.5.6)成立.

现在来证式(5.5.6). 从式(3.5.7)知, 可取 $q'\in(q,\ 1)$, 当 t 充分大时, 对任意 $\delta\in(0,\ 1)$ 有

$$\frac{a_t}{a_{(1+\delta)t}}\leqslant 1+q'\delta. \tag{5.5.8}$$

设 $t>s$. 我们可定义 δ_t, 使 $t=(1+\delta_t)s$. 那么

$$\frac{a_s}{a_t}=\frac{a_s}{a_{(1+\delta_t)s}}\leqslant 1+q'\delta_t,$$

或

$$\frac{a_s-a_t}{a_s a_t}\leqslant\frac{q'\delta_t}{a_s}=\frac{q'(t-s)}{s a_s}.$$

所以为证式(5.5.6), 只要证明 $t a_t\xrightarrow[t\to\infty]{}\infty$.

固定 $\delta\in(0,\ 1)$ 及 s. 记 $\mu_t=\dfrac{1}{t a_t}$. 设 n 充分大, 使对 $t\geqslant(1+(n-1)\delta)s$ 时, 式(5.5.8)成立. 那么就有

$$\mu_{(1+n\delta)s}=\frac{a_{(1+(n-1)\delta)s}}{a_{(1+(n-1)\delta)s}\,a_{(1+n\delta)s}\cdot s(1+(n-1)\delta)\cdot\dfrac{1+n\delta}{1+(n-1)\delta}}$$

$$\leqslant\mu_{(1+(n-1)\delta)s}\cdot\left(1+\frac{q'\delta}{1+(n-1)\delta}\right)\left(\frac{1+(n-1)\delta}{1+n\delta}\right).$$

设 $n\geqslant n_0$, n_0 充分大, 使 $\dfrac{1+(n-1)\delta}{1+n\delta}>p>q'$, p 为常数, 那么从上式知

$$\mu_{(1+n\delta)s}\leqslant\left(1+\frac{q'\delta}{1+(n-1)\delta}\right)\left(1-\frac{\delta}{1+n\delta}\right)\mu_{(1+(n-1)\delta)s}$$

$$\leqslant\left(1-\frac{\delta}{1+n\delta}+\frac{q'\delta}{1+(n-1)\delta}\right)\mu_{(1+(n-1)\delta)s}$$

$$<\left(1-\frac{(p-q')\delta}{1+(n-1)\delta}\right)\mu_{(1+(n-1)\delta)s}<\cdots$$

$$\leqslant \prod_{i=n_0}^{n-1}\left(1-\frac{(p-q')\delta}{1+i\delta}\right)\mu_{(1+n_0\delta)s}.$$

因为 $\sum\limits_{i=n_0}^{\infty}\frac{(p-q')\delta}{1+i\delta}=\infty$, 我们就有

$$\prod_{i=n_0}^{n}\left(1-\frac{(p-q')\delta}{1+i\delta}\right)\xrightarrow[n\to\infty]{}0,$$

于是 $\mu_{(1+n\delta)s}\xrightarrow[n\to\infty]{}0$.

显然, 对任意 t, 存在 n 及 $x\in[0,\delta)$ 使 $t=(1+(n-1)\delta+x)s$. 注意到

$$(1+n\delta)s=(1+\delta')(1+(n-1)\delta+x)s,$$

$$\delta'=\frac{\delta-x}{1+(n-1)\delta+x},$$

我们就有对 $x\in[0,\delta)$ 的一致收敛:

$$\begin{aligned}\frac{\mu_{(1+n\delta)s}}{\mu_t}&=\frac{(1+(n-1)\delta+x)s a_{(1+(n-1)\delta+x)s}}{(1+n\delta)s a_{(1+n\delta)s}}\\&=\left(1+\frac{x-\delta}{1+n\delta}\right)\frac{a_{(1+(n-1)\delta+x)s}}{a_{(1+n\delta)s}}\\&\leqslant\left(1+\frac{x-\delta}{1+n\delta}\right)(1+\beta\delta')\xrightarrow[n\to\infty]{}1,\end{aligned}$$

这和 $\mu_{(1+n\delta)s}\xrightarrow[n\to\infty]{}0$ 一起就证明了 $\mu_t\xrightarrow[t\to\infty]{}0$.

引理 5.5.1　设 a_t 为慢衰减增益. 那么 a_t 有以下性质:

i)
$$ta_t\xrightarrow[t\to\infty]{}\infty, \tag{5.5.9}$$

ii) 设 $t\geqslant s$, $o(1)\xrightarrow[s\to\infty]{}0$, 那么

$$\frac{a_s}{a_t}\leqslant\exp\left\{o(1)\int_s^t a_\lambda \mathrm{d}\lambda\right\}, \tag{5.5.10}$$

$$\left|\frac{a_s}{a_t}-1\right|\leqslant o(1)\int_s^t a_\lambda\,\mathrm{d}\lambda\exp\left\{o(1)\int_s^t a_\lambda \mathrm{d}\lambda\right\}, \tag{5.5.11}$$

iii) 对任意 $\delta\in(0,1)$,

$$\int_{t(1-\delta)}^{t} a_s\,\mathrm{d}s\xrightarrow[t\to\infty]{}\infty. \tag{5.5.12}$$

证明　记

$$\mu_t=\frac{1}{ta_t}. \tag{5.5.13}$$

从式(5.5.6)知

$$\mu_t=\frac{s}{t}\mu_s+o(1)\frac{t-s}{t}. \qquad (5.5.14)$$

取 s_0 充分大，使对 $s\geqslant s_0$，$|o(1)|<\varepsilon$. 在式(5.5.14)中固定 $s\geqslant s_0$，并取 $t\to\infty$，便知当 t 充分大时

$$\mu_t<2\varepsilon.$$

这就证明了式(5.5.9).

从式(5.5.6)知

$$\begin{aligned}\frac{a_s}{a_t}&=1+o(1)(t-s)a_s=1+o(1)\int_s^t a_\lambda\frac{a_s}{a_\lambda}\,\mathrm{d}\lambda\\&=1+o(1)\int_s^t a_\lambda\left[1+o(1)\int_s^\lambda a_\tau\frac{a_s}{a_\tau}\,\mathrm{d}\tau\right]\mathrm{d}\lambda\\&=\cdots\leqslant\exp\left\{o(1)\int_s^t a_\lambda\mathrm{d}\lambda\right\},\end{aligned}$$

这就证明了式(5.5.10).

对任一 $\delta\in(0,1)$，由式(5.5.9)就有

$$\begin{aligned}\int_{t(1-\delta)}^t a_s\,\mathrm{d}s&=\int_{t(1-\delta)}^t\frac{sa_s}{s}\,\mathrm{d}s\geqslant k\int_{t(1-\delta)}^t\frac{\mathrm{d}s}{s}\\&=k\log\frac{1}{1-\delta}\xrightarrow[k\to\infty]{}\infty.\end{aligned}$$

□

设 H 为稳定阵，用 $\Phi_{t,s}$ 表示基本解阵

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t,s}=a_tH\Phi_{t,s},\ \Phi_{s,s}=I,\ t\geqslant s\geqslant 0. \qquad (5.5.15)$$

容易验证

$$\Phi_{t,s}=\exp\left\{H\int_s^t a_\lambda\mathrm{d}\lambda\right\}. \qquad (5.5.16)$$

由于 H 稳定，所以存在常数 $c_0>0$, $c>0$ 使

$$\|\Phi_{t,s}\|\leqslant c_0\exp\left\{-c\int_s^t a_\lambda\mathrm{d}\lambda\right\}. \qquad (5.5.17)$$

引理 5.5.2 设 a_t 为慢衰减增益，H 稳定，$\Phi_{t,s}$ 由式(5.5.15)给出，那么

$$a_s\int_s^t\Phi_{\lambda,s}\,\mathrm{d}\lambda,$$

对 s, t 一致有界：$0\leqslant s\leqslant t<\infty$，并且对任意 $\varepsilon>0$,

$$\frac{1}{t}\int_s^t a_s\int_s^t \Phi_{\lambda,s}\,\mathrm{d}\lambda\,\mathrm{d}s\xrightarrow[t\to\infty]{} H^{-1}. \tag{5.5.18}$$

证明 由式(5.5.15)知

$$\Phi_{t,s}=I+H\int_s^t a_\lambda\Phi_{\lambda,s}\,\mathrm{d}\lambda,$$

所以由式(5.5.12)及(5.5.17)就有

$$\int_s^t a_\lambda\Phi_{\lambda,s}\mathrm{d}\lambda=-H^{-1}+H^{-1}\Phi_{t,s}\xrightarrow[t\to\infty]{}-H^{-1}. \tag{5.5.19}$$

记

$$G_{t,s}=\int_s^t (a_s-a_\lambda)\Phi_{\lambda,s}\mathrm{d}\lambda+H^{-1}\Phi_{t,s}, \tag{5.5.20}$$

那么

$$\begin{aligned} a_s\int_s^t \Phi_{\lambda,s}\,\mathrm{d}\lambda&=\int_s^t (a_s-a_\lambda)\Phi_{\lambda,s}\mathrm{d}\lambda+\int_s^t a_\lambda\Phi_{\lambda,s}\,\mathrm{d}\lambda\\ &=-H^{-1}+G_{t,s}. \end{aligned} \tag{5.5.21}$$

为证引理,只要证明 $G_{t,s}$ 对 $0\leqslant s\leqslant t<\infty$ 一致有界,并且

$$\frac{1}{t}\int_s^t \|G_{t,s}\|\mathrm{d}s\xrightarrow[t\to\infty]{}0. \tag{5.5.22}$$

由式(5.5.12)和(5.5.17)知,对任一 $\delta\in(0,1)$,

$$\begin{aligned} \frac{1}{t}\int_s^t\|\Phi_{t,s}\|\mathrm{d}s&\leqslant\frac{c_0}{t}\left(\int_s^{(1-\delta)t}+\int_{(1-\delta)t}^t\right)\exp\left\{-c\int_s^t a_\lambda\mathrm{d}\lambda\right\}\mathrm{d}s\\ &\leqslant\frac{c_0}{t}\int_s^{(1-\delta)t}\exp\left\{-c\int_s^t a_\lambda\mathrm{d}\lambda\right\}\mathrm{d}s+c_0\delta\\ &\leqslant\frac{c_0(1-\delta)t}{t}\exp\left\{-c\int_s^t a_\lambda\mathrm{d}\lambda\right\}\mathrm{d}s+c_0\delta\\ &\leqslant\frac{c_0(1-\delta)t}{t}\exp\left\{-c\int_{(1-\delta)t}^t a_\lambda\mathrm{d}\lambda\right\}\\ &\quad+c_0\delta\xrightarrow[\substack{t\to\infty\\ \delta\to 0}]{}0. \end{aligned} \tag{5.5.23}$$

所以为证式(5.5.22),只要证明

$$\frac{1}{t}\int_s^t\int_s^t (a_s-a_\lambda)\Phi_{\lambda,s}\,\mathrm{d}\lambda\,\mathrm{d}s\xrightarrow[t\to\infty]{}0. \tag{5.5.24}$$

简记
$$g_{t,s}=\int_s^t a_\lambda \mathrm{d}\lambda,$$
并认为 s_0 已充分大，使对 $s>s_0$ 有 $o(1)-c<-c_1$，$c_1>0$. 由式(5.5.17)知

$$\begin{aligned}
&\left\|\frac{1}{t}\int_s^t\int_s^t (a_s-a_\lambda)\Phi_{\lambda,s}\mathrm{d}\lambda\,\mathrm{d}s\right\| \\
&\quad\leqslant\frac{c_0}{t}\int_s^t\int_s^t\left(\frac{a_s-a_\lambda}{a_\lambda}\right)[\exp(-cg_{\lambda,s})]a_\lambda\mathrm{d}\lambda\,\mathrm{d}s \\
&\quad=\frac{c_0}{t}\left(\int_s^{s_0}+\int_{s_0}^t\right)\int_s^t\left(\frac{a_s}{a_\lambda}-1\right)[\exp(-cg_{\lambda,s})]a_\lambda\mathrm{d}\lambda\,\mathrm{d}s.
\end{aligned}\tag{5.5.25}$$

由式(5.5.10)知

$$\begin{aligned}
&\left|\frac{c_0}{t}\int_{s_0}^t\int_s^t\left(\frac{a_s-a_\lambda}{a_\lambda}\right)[\exp(-cg_{\lambda,s})]a_\lambda\mathrm{d}\lambda\,\mathrm{d}s\right| \\
&\quad\leqslant\frac{c_0}{t}\int_{s_0}^t o(1)\int_s^t g_{\lambda,s}\exp((o(1)-c)g_{\lambda,s})\,\mathrm{d}g_{\lambda,s}\,\mathrm{d}s \\
&\quad<\frac{c_0}{t}\int_{s_0}^t o(1)\int_s^t g_{\lambda,s}\exp(-c_1g_{\lambda,s})\mathrm{d}f_{\lambda,s}\,\mathrm{d}s \\
&\quad=\frac{c_0}{t}\int_{s_0}^t o(1)\int_0^{g_{t,s}} y\exp(-c_1y)\mathrm{d}y\,\mathrm{d}s\xrightarrow[s_0\to\infty]{}0.
\end{aligned}\tag{5.5.26}$$

在式(5.5.25)中的另一个积分可估计如下：

$$\begin{aligned}
&\frac{c_0}{t}\int_s^{s_0}\int_s^t(a_s-a_\lambda)\exp(-cg_{\lambda,s})\,\mathrm{d}\lambda\,\mathrm{d}s \\
&\quad=\frac{c_0}{t}\int_s^{s_0}\left(\int_s^{\delta t}+\int_{\delta t}^t\right)(a_s-a_\lambda)\exp(-cg_{\lambda,s})\mathrm{d}\lambda\,\mathrm{d}s \\
&\quad\leqslant\frac{c_0a_s}{t}[s_0(\delta t-s)+(1-\delta)ts_0\exp(-cg_{\delta t,s_0})]\xrightarrow[\substack{t\to\infty\\ \delta\to 0}]{}0.
\end{aligned}\tag{5.5.27}$$

式(5.5.25)～(5.5.27)一起证明了式(5.5.24).

在式(5.5.26)中，我们还证明了 $\int_s^t(a_s-a_\lambda)\Phi_{\lambda,s}\mathrm{d}\lambda$ 对 s，t 一致有界，$t\geqslant s\geqslant s_0$. 从式(5.5.17)、(5.5.20)和(5.5.21)还知 $G_{t,s}$

及 $a_s\int_s^t\Phi_{\lambda,s}\mathrm{d}\lambda$ 在 $t\geqslant s\geqslant s_0$ 上有界. 而对 $0\leqslant s<s_0$, 它们的有界性可从下面的估计看出:

$$\begin{aligned}&\left\|a_s\int_s^t\Phi_{\lambda,s}\mathrm{d}\lambda\right\|\\&=\left\|a_s\int_s^{s_0}\Phi_{\lambda,s}\mathrm{d}\lambda+\frac{a_s}{a_{s_0}}a_{s_0}\int_{s_0}^t\Phi_{\lambda,\lambda_0}\mathrm{d}\lambda\,\Phi_{s_0,s}\right\|\\&\leqslant a_sc_0s_0+\frac{a_sc_0}{a_{s_0}}\left\|a_{s_0}\int_{s_0}^t\Phi_{\lambda,s_0}\mathrm{d}\lambda\right\|.\end{aligned}$$ □

下面我们要给出本节的主要内容. 为此, 我们先引进条件.

A 5.5.0 $\|F(\cdot)\|$有界, $f=\sup\limits_{x\in R^l}\|F(x)\|$, $h(\cdot)$及 $F(\cdot)$满足局部 Lipschitz 条件, $h(x^0)=0$, 存在稳定阵 H 及实数 $\gamma>0$, $\beta\in(0,1]$, $k>0$, 使

$$\begin{aligned}&\|h(x)-H(x-x^0)\|\\&\leqslant k\|x-x^0\|^{1+\beta},\ \forall x\in\{x:\|x-x^0\|\leqslant\gamma\}.\end{aligned}\tag{5.5.28}$$

A 5.5.1 $\{a_t\}$ 为慢衰减增益, 并且对某个 $\alpha\leqslant\beta$, $\alpha\in(0,1)$

$$\int_s^\infty\left(\frac{a_s^{1+\alpha}}{s}\right)^{\frac12}\mathrm{d}s<\infty,\ \forall s>0.\tag{5.5.29}$$

A 5.5.2 存在二次连续可微函数 $v(\cdot)$: $R^l\to R$, 使

$$v(x^0)=0,\ v(x)\neq0,\ h^\tau(x)v_x(x)<0,\ \forall x\neq x^0,\tag{5.5.30}$$

并且对固定的 x^*, 存在常数 $M>0$, $\|x^*\|<M$, 使

$$v(x^*)<\inf_{\|x\|=M}v(x).$$

A 5.5.3 $\mathscr{F}_t\uparrow$, 右连续, $(w_t,\mathscr{F}_t)$为 l 维 Wiener 过程, η_t 和 $\mathscr{F}_t$ 适应, 并且

$$\eta_t\xrightarrow[t\to\infty]{}0,\qquad \mathrm{E}\|\eta_t\|^2=O(a_t^{1+\alpha}).\tag{5.5.31}$$

从式(5.5.29)立刻看出

$$\int_{\frac t2}^t\left(\frac{a_s^{1+\alpha}}{s}\right)^{\frac12}\mathrm{d}s\geqslant\frac t2\left(\frac{a_t^{1+\alpha}}{t}\right)^{\frac12},$$

所以

$$a_t = o(t^{-\mu}), \quad \mu = \frac{1}{1+\alpha} \in \left(\frac{1}{2}, 1\right), \tag{5.5.32}$$

因此

$$\int_s^\infty a_s^2 \, ds < \infty. \tag{5.5.33}$$

由此可见，当 A5.5.0～A5.5.3 成立时，定理 5.2.1 的条件全部成立. 所以由式(5.2.11)定义的 x_t，当 $t\to\infty$ 时 a.s. 收敛到 x^0. 同时存在依赖于样本的 i_0，使 $\alpha_{i_0}<\infty$ a.s.，但 $\alpha_{i_0+1}=\infty$ a.s.，并且

$$dx_t = a_t h(x_t) dt + a_t d\varepsilon_t, \ \forall t \geqslant \alpha_{i_0}. \tag{5.5.34}$$

根据式(5.5.15)，从式(5.5.34)便知

$$x_t - x^0 = \Phi_{t,\alpha_{i_0}}(x_{\alpha_{i_0}} - x^0) + \int_{\alpha_{i_0}}^t \Phi_{t,s} a_s [h(x_s) - H(x_s - x^0) + \eta_s] ds$$
$$+ \int_{\alpha_{i_0}}^t \Phi_{t,s} a_s F(x_s) dw_s \quad \forall t \geqslant \alpha_{i_0}, \tag{5.5.35}$$

对式(5.2.11)定义的 x_t，固定 $\gamma\in(0, 1)$及确定性时间 t_0 定义停时 μ:

$$\mu = \begin{cases} \inf\{t: t > t_0, \ \|x_t - x^0\| > \gamma,\} \\ 0, \quad \text{如果 } \|x_{t_0} - x^0\| > \gamma. \end{cases} \tag{5.5.36}$$

由式(5.5.35)知

$$\begin{aligned}
&(x_t - x^0) I_{[\mu>t, \alpha_{i_0}<t_0]} \\
&= \Phi_{t,t_0}(x_{t_0} - x^0) I_{[\mu>t, \alpha_{i_0}<t_0]} \\
&\quad + \int_{t_0}^t \Phi_{t,s} a_s [h(x_s) - H(x_s - x^0) + \eta_s] ds I_{[\mu>t, \alpha_{i_0}<t_0]} \\
&\quad + \int_{t_0}^t \Phi_{t,s} a_s F(x_s) dw_s I_{[\mu>t, \alpha_{i_0}<t_0]}, \quad \forall t \geqslant t_0.
\end{aligned} \tag{5.5.37}$$

引理 5.5.3 设条件 A 5.5.0～A 5.5.3 成立，那么对任意 $\gamma>0$, $t_0>0$,

$$\frac{1}{a_t} \mathrm{E}(\|x_t - x^0\|^2 I_{[\mu>t, \alpha_{i_0}<t_0]})$$

对 $t\geqslant t_0$ 有界.

证明 根据式(5.5.17)及式(5.4.1)，从式(5.5.37)知

$$\frac{1}{a_t} \mathrm{E}(\|x_t - x^0\|^2 I_{[\mu>t, \alpha_{i_0}<t_0]})$$

$$
\begin{aligned}
&\leqslant \frac{4c_0^2\gamma^2}{a_t}\exp\left(-2cg_{t,t_0}\right) \\
&\quad+\frac{4c_0^2}{a_t}\mathrm{E}\left(\int_{t_0}^{t}\left[\exp(-cg_{t,s})\right]a_s c_1\|x_s-x^0\|^{1+\beta}\right. \\
&\quad\left.\cdot I_{[\mu>s,\,a_{t_0}<t_0]}\mathrm{d}s\right)^2 \\
&\quad+\frac{4c_0^2}{a_t}\mathrm{E}\left(\int_{t_0}^{t}a_s\|\eta_s\|\exp\{-g_{t,s}\}\mathrm{d}s\right)^2 \\
&\quad+\frac{4c_0^2}{a_t}\int_{t_0}^{t}\left[\exp(-2cg_{t,s})\right]a_s^2\mathrm{E}\|F(x_s)\|^2 I_{[\mu>s,\,a_{t_0}<t_0]}\mathrm{d}s.
\end{aligned}
\tag{5.5.38}
$$

把上式右端的四项分别用 I_i, $i=1, \cdots, 4$ 来表示，然后来分别估计它们.

根据式(5.5.10)，当 t_0 充分大时有

$$
I_1\leqslant\frac{4c_0^2\gamma^2}{a_{t_0}}\cdot\frac{a_{t_0}}{a_t}\exp\left(-2cg_{t,t_0}\right)\leqslant\frac{4c_0^2\gamma^2}{a_{t_0}}\exp(-cg_{t,t_0}),
\tag{5.5.39}
$$

上式右端显然对 t 有界.

从式(5.5.10)及式(5.5.31)可知，当 t_0 充分大时，有

$$
\begin{aligned}
I_3&\leqslant\frac{4c_0^2}{a_t}\int_{t_0}^{t}\left[\exp(-cg_{t,s})\right]a_s^2\,\mathrm{d}s\int_{t_0}^{t}\left[\exp(-cg_{t,s})\right]a_s^{1+\alpha}\mathrm{d}s \\
&\leqslant a_{t_0}^{\alpha}4c_0^2\int_{t_0}^{t}a_s\exp\left\{-\frac{c}{2}g_{t,s}\right\}\mathrm{d}s\int_{t_0}^{t}a_s\exp\{-cg_{t,s}\}\mathrm{d}s,
\end{aligned}
\tag{5.5.40}
$$

$$
I_4\leqslant 4c_0^2\max_{\|x\|\leqslant\gamma+\|x^0\|}\|F(x)\|^2\int_{t_0}^{t}\exp\{-cg_{t,s}\}a_s\,\mathrm{d}s,
\tag{5.5.41}
$$

所以 I_3 及 I_4 两项都对 t 有界.

现在来估 I_2.

$$
\begin{aligned}
I_2\leqslant&\frac{4c_0^2c_1^2}{a_t}\int_{t_0}^{t}a_s^2\exp\{-cg_{t,s}\}\mathrm{d}s \\
&\cdot\int_{t_0}^{t}\left[\exp(-cg_{t,s})\right]\mathrm{E}\|x_s-x^0\|^{2(1+\beta)}\cdot I_{[\mu>s,\,a_{t_0}<t_0]}\mathrm{d}s
\end{aligned}
$$

$$\leqslant 4c_0^2c_1^2\gamma^{2\beta}\int_{t_0}^t a_s\exp\left(-\frac{c}{2}g_{t,s}\right)\mathrm{d}s\int_{t_0}^t\left[\exp(-cg_{t,s})\right]$$

$$\cdot\mathrm{E}\|x_s-x^0\|^2 I_{[\mu>s,\,a_{t_0}<t_0]}\mathrm{d}s. \qquad (5.5.42)$$

综合式(5.5.38)～(5.5.42)，可知对充分大的 t_0，存在常数 $k_0>0$, $k_1>0$ 使

$$\frac{1}{a_t}\mathrm{E}(\|x_t-x^0\|^2 I_{[\mu>t,\,a_{t_0}<t_0]})$$

$$\leqslant k_0+k_1\gamma^{2\beta}\int_{t_0}^t\left[\exp(-cg_{t,s})\right]\mathrm{E}\|x_s-x^0\|^2 I_{[\mu>s,\,a_{t_0}<t_0]}\mathrm{d}s. \qquad (5.5.43)$$

记

$$g_t=\left[\exp(cg_{t,0})\right]\frac{1}{a_t}\mathrm{E}(\|x_t-x^0\|^2 I_{[\mu>t,\,a_{t_0}<t_0]}), \qquad (5.5.44)$$

由式(5.5.43)得:

$$g_t\leqslant k_0\exp(cg_{t,0})+k_1\gamma^{2\beta}\int_{t_0}^t a_s g_s\,\mathrm{d}s.$$

由 Bellman-Gronwall 不等式知

$$g_t\leqslant k_0\exp\{cg_{t,0}\}+k_1\gamma^{2\beta}\int_{t_0}^t a_s k_0\left[\exp(cg_{s,0})\right]\exp(k_1\gamma^{2\beta}g_{t,s})\mathrm{d}s,$$

或等价地

$$\frac{1}{a_t}\mathrm{E}(\|x_t-x^0\|^2 I_{[\mu>t,\,a_{t_0}<t_0]})$$

$$\leqslant k_0+k_0k_1\gamma^{2\beta}\int_{t_0}^t a_s\exp((-c+k_1\gamma^{2\beta})g_{t,s})\,\mathrm{d}s. \qquad (5.5.45)$$

取 γ 充分小，使 $k_1\gamma^{2\beta}\leqslant\frac{c}{2}$. 引理结论直接可由式(5.5.45)得出. □

引理 5.5.4 设 A5.5.0～A5.5.3 成立. 那么

$$\frac{1}{\sqrt{t}}\int_0^t\|h(x_s)-H(x_s-x^0)\|\mathrm{d}s\xrightarrow[t\to\infty]{}0 \quad \text{a.s.}, \qquad (5.5.46)$$

证明 类似于离散时间的 Kronecker 引理，我们有下面结果：如果对某个 $d>0$,

$$a_t \triangleq \int_d^t \frac{b_s}{\sqrt{s}} \mathrm{d}s \xrightarrow[t\to\infty]{} a \triangleq \int_d^\infty \frac{b_s}{\sqrt{s}} \mathrm{d}s<\infty,$$

那么

$$\frac{1}{\sqrt{t}} \int_d^t b_s \mathrm{d}s = \frac{1}{\sqrt{t}} \int_d^t \sqrt{s}\, \mathrm{d}a_s = a_t - \frac{1}{\sqrt{t}} \int_d^t \frac{a_s}{2\sqrt{s}} \mathrm{d}s$$

$$= a_t - a + \frac{a\sqrt{d}}{\sqrt{t}} - \frac{1}{\sqrt{t}} \int_d^t \frac{(a_s - a)}{2\sqrt{s}} \mathrm{d}s \xrightarrow[t\to\infty]{} 0.$$

所以，为了证式(5.5.46)只要证明对某个 $d>0$，有

$$\int_d^\infty \frac{1}{\sqrt{s}} \|h(x_s) - H(x_s - x^0)\| \mathrm{d}s < \infty \quad \text{a.s.}. \tag{5.5.47}$$

利用式(5.5.29)及引理5.5.3知

$$\mathrm{E}\int_{t_0}^\infty \frac{1}{\sqrt{s}} \|h(x_s) - H(x_s - x^0)\| I_{[\mu>s,\, \alpha_{t_0}<t_0]} \mathrm{d}s$$

$$\leqslant \int_{t_0}^\infty \frac{\beta}{\sqrt{s}} \mathrm{E}\|x_s - x^0\|^{1+\beta} I_{[\mu>s,\alpha_{t_0}<t_0]}$$

$$\leqslant \beta k_2 \int_{t_0}^\infty \frac{1}{\sqrt{s}} a_s^{\frac{1+\beta}{2}} \mathrm{d}s<\infty,$$

这里 k_2 表示 $\left(\dfrac{\mathrm{E}\|x_s - x^0\|^2 I_{[\mu>s,\alpha_{t_0}<t_0]}}{a_s}\right)^{\frac{1+\beta}{2}}$ 的上界. 所以

$$\int_{t_0}^\infty \frac{1}{\sqrt{s}} \|h(x_s) - H(x_s - x^0)\| I_{[\mu>s,\alpha_{t_0}<t_0]} \mathrm{d}s < \infty \quad \text{a.s.},$$

因此

$$\left\{\int_{t_0}^\infty \frac{1}{\sqrt{s}} \|h(x_s) - H(x_s - x^0)\| \mathrm{d}s<\infty\right\}$$

$$\supset \{\sup_{t_0\leqslant t<\infty} \|x_t - x^0\| < \gamma,\ \alpha_{t_0}<t_0\}. \tag{5.5.46}$$

据定理5.2.1，$x_t \xrightarrow[t\to\infty]{} x^0$ a.s.，注意到 α_{t_0} 有穷 a.s.，所以对任意 $\varepsilon>0$，只要 t_0 充分大必有

$$P\{\sup_{t_0\leqslant t<\infty} \|x_t - x^0\| < \gamma,\ \alpha_{t_0}<t_0\} > 1-\varepsilon. \tag{5.5.49}$$

由式(5.5.48)、(5.5.49)知

$$P\left\{\int_d^\infty \frac{1}{\sqrt{s}} \|h(x_s) - H(x_s - x^0)\| \mathrm{d}s<\infty\right\} > 1-\varepsilon. \tag{5.5.50}$$

因为 ε 任意，所以由式(5.5.50)知

$$\int_{\delta}^{\infty}\frac{1}{\sqrt{s}}\|h(x_s)-H(x_s-x^0)\|\mathrm{d}s<\infty \quad \text{a.s.},$$

这就证明了引理. □

现在我们来叙述和证明本节的主要内容.

定理 5.5.1 设 A 5.5.0～A 5.5.3 成立，x_t 由式(5.2.11) 给出. 取 x_t 的平均值

$$\bar{x}_t=\frac{1}{t}\int_0^t x_s\,\mathrm{d}s, \tag{5.5.51}$$

那么

$$\sqrt{t}\,(\bar{x}_t-x^0)\xrightarrow[t\to\infty]{d}N(0,\,S), \tag{5.5.52}$$

$$S=H^{-1}F(x^0)F^{\tau}(x^0)H^{-\tau}, \tag{5.5.53}$$

即 $\bar{x}_t$ 是对 x^0 的渐近有效估计.

证明 从式(5.5.35)得

$$\begin{aligned}\bar{x}_t=&\frac{1}{t}\int_0^{\alpha_{i_0}}x_s\,\mathrm{d}s+\frac{t-\alpha_{i_0}}{t}x^0+\frac{1}{t}\int_{\alpha_{i_0}}^t\Phi_{\lambda,\alpha_{i_0}}x_{\alpha_{i_0}}\,\mathrm{d}\lambda\\&+\frac{1}{t}\int_{\alpha_{i_0}}^t\int_{\alpha_{i_0}}^{\lambda}a_s\Phi_{\lambda,s}[(h(x_s)-H(x_s-x^0))\mathrm{d}s\\&+\eta_s\mathrm{d}s+F(x_s)\mathrm{d}w_s]\mathrm{d}\lambda,\end{aligned}$$

以及

$$\begin{aligned}&\sqrt{t}\,(\bar{x}_t-x^0)\\&=-\frac{\alpha_{i_0}}{\sqrt{t}}x^0+\frac{1}{\sqrt{t}}\int_0^{\alpha_{i_0}}x_s\,\mathrm{d}s+\frac{1}{\sqrt{t}}\int_{\alpha_{i_0}}^t\Phi_{\lambda,\alpha_{i_0}}\mathrm{d}\lambda x_{\alpha_{i_0}}\\&\quad+\frac{1}{\sqrt{t}}\int_{\alpha_{i_0}}^t\left(\int_s^t a_s\Phi_{\lambda,s}\,\mathrm{d}\lambda\right)[h(x_s)-H(x_s-x^0)]\,\mathrm{d}s\\&\quad+\frac{1}{\sqrt{t}}\int_{\alpha_{i_0}}^t\int_s^t a_s\Phi_{\lambda,s}\,\mathrm{d}\lambda\,\eta_s\,\mathrm{d}s\\&\quad+\frac{1}{\sqrt{t}}\int_{\alpha_{i_0}}^t\left(\int_s^t a_s\Phi_{\lambda,s}\,\mathrm{d}\lambda\right)F(x_s)\mathrm{d}w_s\\&\triangleq\sum_{i=1}^{6}I_i,\end{aligned} \tag{5.5.54}$$

这里用 I_i 表示 $\sqrt{t}\,(\bar{x}_t-x^0)$ 的表达式中依次的六项.

很显然，$I_1 \xrightarrow[t\to\infty]{} 0$，$I_2 \xrightarrow[t\to\infty]{} 0$ a.s.

由式(5.5.10)、(5.5.17)知

$$\|I_3\| \leqslant \frac{1}{\sqrt{t}} \left\| \int_{a_{t_0}}^{t} \Phi_{\lambda, a_{t_0}} \mathrm{d}\lambda x_{a_{t_0}} \right\|$$

$$+ \frac{c_0}{a_{t_0}\sqrt{t}} \int_{t_0}^{t} \frac{a_{t_0}}{a_\lambda} a_\lambda \exp(-cg_{\lambda, t_0}) \mathrm{d}\lambda \exp(-cg_{t_0, a_{t_0}}),$$

当 t_0 充分大时，从上式便得

$$\|I_3\| \leqslant \frac{c_0}{\sqrt{t}} \int_{a_{t_0}}^{t_0} \exp(-cg_{\lambda, a_{t_0}}) \mathrm{d}\lambda \|x_{a_{t_0}}\|$$

$$+ \frac{c_0}{a_{t_0}\sqrt{t}} \int_{t_0}^{t} a_\lambda \exp\left(-\frac{c}{2} g_{\lambda, t_0}\right) \mathrm{d}\lambda \exp(-cg_{t_0 a_{t_0}})$$

$$\xrightarrow[t\to\infty]{} 0 \quad \text{a.s.}, \tag{5.5.55}$$

由式(5.5.21)和引理5.5.4以及 $G_{t,s}$ 的有界性(引理5.5.2)就有

$$\|I_4\| \leqslant H^{-1} \frac{1}{\sqrt{t}} \left\| \int_{a_{t_0}}^{t} [h(x_s) - H(x_s - x^0)] \mathrm{d}s \right\|$$

$$+ \frac{1}{\sqrt{t}} \left\| \int_{a_{t_0}}^{t} G_{t,s} [h(x_s) - H(x_s - x^0)] \mathrm{d}s \right\|$$

$$\xrightarrow[t\to\infty]{} 0 \quad \text{a.s.}, \tag{5.5.56}$$

再由式(5.5.29)、(5.5.31)及 $G_{t,s}$ 的有界性便知

$$\mathrm{E} \int_1^\infty \frac{\|G_{t,s}\eta_s\|}{\sqrt{s}} \mathrm{d}s \leqslant c_2 \int_1^\infty \frac{a_s^{\frac{1+\alpha}{2}}}{\sqrt{s}} \mathrm{d}s < \infty,$$

因此
$$\frac{1}{\sqrt{t}} \int_1^t G_{t,s} \eta_s \mathrm{d}s \xrightarrow[t\to\infty]{} 0 \quad \text{a.s.},$$

所以
$$\frac{1}{\sqrt{t}} \int_{a_{t_0}}^{t} G_{t,s} \eta_s \mathrm{d}s \xrightarrow[t\to\infty]{} 0 \quad \text{a.s.},$$

即

$$I_5 \xrightarrow[t\to\infty]{} 0 \quad \text{a.s.}, \tag{5.5.57}$$

现在来估 I_6.

$$I_6=\frac{-H^{-1}}{\sqrt{t}}\int_{a_{t_0}}^{t}F(x_s)\mathrm{d}w_s+\frac{1}{\sqrt{t}}\int_{a_{t_0}}^{t}G_{t,s}F(x_s)\mathrm{d}w_s$$

$$=-\frac{H^{-1}}{\sqrt{t}}F(x^0)w_t+\frac{H^{-1}}{\sqrt{t}}F(x^0)w_{a_{t_0}}$$

$$-\frac{H^{-1}}{\sqrt{t}}\int_{a_{t_0}}^{t}(F(x_s)-F(x^0))\mathrm{d}w_s$$

$$+\frac{1}{\sqrt{t}}\int_{a_{t_0}}^{t}G_{t,s}F(x_s)\mathrm{d}w_s.$$

由于 $x_s\xrightarrow[s\to\infty]{}x^0$ a.s.，用控制收敛定理知，对任一 $c>0$，

$$\mathrm{E}\|(F(x_s)-F(x^0))I_{[\|x_s-x^0\|<c]}\|^2\xrightarrow[s\to\infty]{}0,$$

因此

$$\frac{1}{t}\mathrm{E}\left\|\int_0^t(F(x_s)-F(x^0))I_{[\|x_s-x^0\|<c]}\mathrm{d}w_s\right\|^2$$

$$=\frac{1}{t}\int_0^t\mathrm{E}\|(F(x_s)-F(x^0))\|^2I_{[\|x_s-x^0\|<c]}\mathrm{d}s\xrightarrow[t\to\infty]{}0,$$

这表明对任一 $c>0$，

$$\frac{1}{\sqrt{t}}\int_{a_{t_0}}^{t}(F(x_s)-F(x^0))I_{[\|x_s-x^0\|<c]}\mathrm{d}w_s\xrightarrow[t\to\infty]{p}0.$$

所以对任一 $\varepsilon>0$，

$$P\left[\left\|\frac{H^{-1}}{\sqrt{t}}\int_{a_{t_0}}^{t}(F(x_s)-F(x^0))\mathrm{d}w_s\right\|>\varepsilon\right]$$

$$\leqslant P\left[\left\|\frac{H^{-1}}{\sqrt{t}}\int_{a_{t_0}}^{t}(F(x_s)-F(x^0))I_{[\|x_s-x^0\|<c]}\mathrm{d}w_s\right\|>\varepsilon\right]$$

$$+P\,[\sup_t\|x_t-x^0\|>c]\xrightarrow[\substack{t\to\infty\\c\to\infty}]{}0. \tag{5.5.58}$$

类似地，由 $G_{t,s}$ 的有界性及式(5.5.22)知

$$P\left[\frac{1}{\sqrt{t}}\left\|\int_{a_{t_0}}^{t}G_{t,s}F(x_s)\mathrm{d}w_s\right\|>\varepsilon\right]$$

$$\leqslant P\left[\frac{1}{\sqrt{t}}\left\|\int_{a_{t_0}}^{t}G_{t,s}F(x_s)I_{[\|x_s\|<c]}\mathrm{d}w_s\right\|>\varepsilon\right]$$

$$+P\,[\sup_t\|x_t\|>c]\xrightarrow[\substack{t\to\infty\\c\to\infty}]{}0. \tag{5.5.59}$$

所以，综合式(5.5.54)～(5.5.59)便知存在$\zeta_t \xrightarrow[t\to\infty]{p} 0$，使

$$\sqrt{t}\,(\bar{x}_t - x^0) = -\frac{H^{-1}}{\sqrt{t}}\,F(x^0)w_t + \zeta_t.$$

由于w_t正态，从上式立即可知式(5.5.52)和式(5.5.53)成立. □

第 6 章

系统参数估计

在系统辨识、逆应控制、离散事件系统、神经元网络以及模式识别等领域中，经常需要分析不同类型递推算法的收敛性，这一章我们将应用前几章给出的随机逼近的收敛性定理，来解决系统参数估计问题．虽然这里只限于分析有限的几个递推算法，但从中我们可以看到如何应用这些收敛定理，以及应用时的难点所在．这一章内容基于参考文献[9, 10, 11]．

§6.1 连续时间动力系统中的参数估计

考察带参数 $\theta\in\Theta$, $\theta\in R^m$ 的动力系统

$$\frac{\mathrm{d}}{\mathrm{d}t}x_t(\theta)=f(\theta,\ x_t(\theta)),\ x_t(\theta)\in R^l. \tag{6.1.1}$$

设对任意终端时间 t_f，我们可找到近似解 $\hat{x}_t(\theta)$，使

$$\|x_t(\theta)-\hat{x}_t(\theta)\|<\varepsilon,\ \forall t\in[0,\ t_f],\ \forall\theta\in\Theta, \tag{6.1.2}$$

$\varepsilon\geqslant0$．当 $\varepsilon=0$ 时，$\hat{x}_t(\theta)$ 就是精确解．

还设 θ^0 是系统中的实际参数，对系统的量测量是

$$y_t=y_0+\int_0^t g_s(x_s(\theta^0))\mathrm{d}s+\varepsilon_t, \tag{6.1.3}$$

ε_t 是量测噪声．

取数列 $\delta_k\xrightarrow[k\to\infty]{}0$, $\delta_k>0$，记 $t_k=k\Delta$, $t_{k+1}-t_k=\Delta>0$．根据式(6.1.2)，可找到近似解 $x_t^k(\theta)$，使

$$\|x_t(\theta)-x_t^k(\theta)\|<\delta_k,\ \forall t\in[0,\ t_{k+1}],\ \forall\theta\in\Theta. \tag{6.1.4}$$

设 $\theta^*\in\Theta$ 是固定点，取数列 $M_i\xrightarrow[i\to\infty]{}\infty$，$M_i>0$. 类似于算法(3.1.3)、(3.1.4)，用变界截尾算法来估计 θ^0:

$$\theta'_k=\theta_k+a_k\left(y_{t_{k+1}}-y_{t_k}-\int_{t_k}^{t_{k+1}}g_s(x_s^k(\theta_k))\mathrm{d}s\right),\qquad(6.1.5)$$

$$\sigma_k=\sum_{i=0}^{k-1}I_{[\|\theta'_i\|>M_{\sigma_i}]},\qquad\sigma_0=0,\qquad(6.1.6)$$

$$\theta_{k+1}=\theta'_kI_{[\|\theta'_k\|\leqslant M_{\sigma_k}]}+\theta^*I_{[\|\theta'_k\|>M_{\sigma_k}]}.\qquad(6.1.7)$$

记

$$h_k(\theta)=\int_{t_k}^{t_{k+1}}[g_s(x_s(\theta^0))-g_s(x_s(\theta))]\mathrm{d}s.\qquad(6.1.8)$$

定理 6.1.1[9] 设成立以下条件.

A 6.1.1 对任意 $\theta\in\Theta$，方程(6.1.1)对任意初值有大范围解，并且成立式(6.1.2)，$\forall s>0$，$g_s(\theta)$ 对 s 一致地对 θ 连续. 对任一 θ 的有界集 $A\subset\Theta$，$h_n(\theta)$ 在 A 上对 n 一致有界;

A 6.1.2 $$a_k>0,\ a_k\to0,\ \sum_{k=1}^{\infty}a_k=\infty;\qquad(6.1.9)$$

A 6.1.3 记

$$m(k,\ t)=\max\left\{m:\ \sum_{i=k}^{m}a_i\leqslant t\right\},\ e_{i+1}=\varepsilon_{t_{i+1}}-\varepsilon_{t_i},$$

当 θ_{n_k} 收敛时，要求

$$\lim_{T\to0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{i=n_k}^{m(n_k,T_k)}a_ie_{i+1}\right\|=0,\ \forall T_k\in[0,\ T];\qquad(6.1.10)$$

A 6.1.4 存在 $R^m\to R$ 的二次可微函数 $v(\cdot)$，它满足以下条件

$$v(\theta^0)=0,\ v(\theta)\neq0,\ \forall\theta\neq\theta^0,\qquad(6.1.11)$$

存在常数 c_0，使 $\|\theta^0\|<c_0$，并且 $v(\theta^*)<\inf\limits_{\|\theta\|=c_0}v(\theta)$，还对任意的 $0<\Delta_1<\Delta_2$，

$$\sup_{\Delta_1<\|\theta-\theta^0\|<\Delta_2}h_n^\tau(\theta)v_\theta(\theta)\leqslant-a,\ a>0,\qquad(6.1.12)$$

这里 $v_\theta(\theta)$ 表示 $v(\theta)$ 的梯度，a 可依赖于 Δ_1 及 Δ_2，但不依赖于 n.

那么，对式(6.1.10)成立的样本，必有

$$\theta_k \xrightarrow[k\to\infty]{} \theta^0.$$

证明 注意到

$$y_{t_{k+1}} - y_{t_k} - \int_{t_k}^{t_{k+1}} g_s(x_s^k(\theta_k))\,\mathrm{d}s$$

$$= \int_{t_k}^{t_{k+1}} [g_s(x_s(\theta^0)) - g_s(x_s(\theta_k))]\,\mathrm{d}s$$

$$+ \int_{t_k}^{t_{k+1}} [g_s(x_s(\theta_k)) - g_s(x_s^k(\theta_k))]\,\mathrm{d}s + \varepsilon_{t_{k+1}} - \varepsilon_{t_k}$$

$$= h_k(\theta_k) + \int_{t_k}^{t_{k+1}} [g_s(x_s(\theta_k)) - g_s(x_s^k(\theta_k))]\,\mathrm{d}s + e_{k+1}.$$

由式(6.1.4)及 $g_s(\theta)$ 对 s 关于 θ 一致地连续，便知

$$\int_{t_k}^{t_{k+1}} [g_s(x_s(\theta_k)) - g_s(x_s^k(\theta_k))]\,\mathrm{d}s \xrightarrow[k\to\infty]{} 0.$$

因此

$$e'_{k+1} \triangleq \int_{t_k}^{t_{k+1}} [g_s(x_s(\theta_k)) - g_s(x_s^k(\theta_k))]\,\mathrm{d}s + e_{k+1} \tag{6.1.13}$$

和 e_{k+1} 一样地满足条件 A6.1.3.

算法(6.1.5)～(6.1.7)可改写成

$$\theta'_k = \theta_k + a_k(h_k(\theta_k) + e'_{k+1}), \tag{6.1.14}$$

$$\sigma_k = \sum_{i=0}^{k-1} I_{[\|\theta'_i\| > M_{\sigma_i}]},\ \sigma_0 = 0, \tag{6.1.15}$$

$$\theta_{k+1} = \theta'_k I_{[\|\theta'_k\| \le M_{\sigma_k}]} + \theta^* I_{[\|\theta'_k\| > M_{\sigma_k}]}. \tag{6.1.16}$$

与定理3.1.1所讨论的算法和条件相对照，唯一的差别在于那里的 $h(\cdot)$ 换成这里时变函数 $h_k(\cdot)$. 但我们要求了式(6.1.12)对 n 一致成立，所以引理 2.4.1 对 $\{\theta_n\}$ 仍成立，引理 3.1.1、3.1.2 以及定理 3.1.1 也照样成立，所以 $\theta_k \xrightarrow[k\to\infty]{} \theta^0$. □

下面我们给出例子，使 A 6.1.1、A 6.1.3 及 A6.1.4 同时成立.

例 1 设式(6.1.1)为线性方程

$$\dot{x}(t) = A(t)x(t), \tag{6.1.17}$$

θ 表示方程的初值，即 $x(0) = \theta$. 记 $\Phi_{t,s}$ 为转移矩阵

$$\dot{\Phi}_{t,0}=A(t)\Phi_{t,0},\ \Phi_{0,0}=I. \tag{6.1.18}$$

对线性方程(6.1.17)很容易求得精确解

$$x_t(\theta)=\Phi_{t,0}\theta, \tag{6.1.19}$$

所以可以在式(6.1.2)中取 $\varepsilon=0$, 在式(6.1.5)中用 $x_s(\theta_k)$ 来取代 $x_s^k(\theta_k)$.

设在式(6.1.3)中 $g_s(x)\equiv x$, 那么

$$h_k(\theta)=\int_{t_k}^{t_{k+1}}\Phi_{s,0}(\theta^0-\theta)\,\mathrm{d}s.$$

设 $\Phi_{t,0}>\delta I$, $\delta>0$, $\forall t\in(0,\infty)$, $\|\Phi_{t,0}\|$ 对 t 一致有界. 在一维($l=1$)时, $\Phi_{t,0}=\exp\left(\int_0^t A(s)\mathrm{d}s\right)$, 这时, 当 $A(s)>0$, $\int_0^\infty A(s)\,\mathrm{d}s<\infty$ 时, 对 $\Phi_{t,0}$ 的要求全部满足. 所以当 $\Phi_{t,0}$ 满足上述条件时, A 6.1.1 成立.

取

$$v(\theta)=\|\theta^0-\theta\|^2,$$

那么

$$\begin{aligned}h_n^\tau(\theta)v_\theta(\theta)&=(\theta^0-\theta)^\tau\int_{t_n}^{t_{n+1}}\Phi_{t_s}^\tau\mathrm{d}s(-2(\theta^0-\theta))\\&\leqslant-2\delta\Delta\|\theta^0-\theta\|^2,\end{aligned}$$

所以 A 6.1.4 成立.

现在来看满足 A 6.1.3 的量测噪声. 设 $(w_t,\ \mathscr{F}_t)$ 为 m 维 Wiener 过程, η_t 为任意过程, $\eta_t\xrightarrow[t\to\infty]{}0$. 当

$$\varepsilon_t=\int_{t_0}^t F_s\mathrm{d}w_s+\int_{t_0}^t\eta_s\,\mathrm{d}s, \tag{6.1.20}$$

并且 $\sum_{k=1}^\infty a_k^2<\infty$, $(F_s,\ \mathscr{F}_s)$为适应过程, $\|F_s\|^2\leqslant f<\infty$, $\forall s\in[t_0,\infty)$时, A6.1.3 也得到满足. 因为这时

$$\left(\int_{t_k}^{t_{k+1}}F_s\mathrm{d}w_s,\ \mathscr{F}_{t_{k+1}}\right)$$

是个鞅差序列, 根据定理 1.2.3, $\sum_{i=1}^\infty a_i\int_{t_i}^{t_{i+1}}F_s\mathrm{d}w_s<\infty$ a.s., 所以式(6.1.10)成立.

例2 考察一维线性系统

$$\dot{x}(t)=\theta A(t)x(t), \quad \int_0^{\infty} A(t)\mathrm{d}t<\infty,\ A(t)>0,$$
$$x(0)\neq 0, \tag{6.1.21}$$

为固定起见，设 $\theta^0>0$, $x(0)>0$.

在式(6.1.5)中仍不必用近似解 $x_s^k(\theta)$，而用精确解

$$x_s(\theta)=x(0)\exp\left\{\theta\int_0^t A(s)\,\mathrm{d}s\right\}. \tag{6.1.22}$$

这时，

$$h_k(\theta)=\int_{t_k}^{t_{k+1}}\left(\exp\left(\theta^0\int_0^t A(s)\,\mathrm{d}s\right)\right.$$
$$\left.-\exp\left(\theta\int_0^t A(s)\,\mathrm{d}s\right)\right)\mathrm{d}t\, x(0),$$

取 $$v(\theta)=\left(\exp\left(\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)-\exp\left(\theta\int_0^{\infty}A(s)\,\mathrm{d}s\right)\right)^2,$$

那么

$$\begin{aligned}
&v_\theta(\theta)h_k(\theta)\\
&=-2\int_0^{\infty}A(s)\mathrm{d}s\left[\exp\left(\theta\int_0^{\infty}A(s)\mathrm{d}s\right)\right]\\
&\quad\cdot\left(\exp\left(\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)-\exp\left(\theta\int_0^{\infty}A(s)\mathrm{d}s\right)\right)\\
&\quad\cdot\left\{\left(\exp\left(\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)-\exp\left(\theta\int_0^{\infty}A(s)\mathrm{d}s\right)\right)\Delta\right.\\
&\quad+\int_{t_k}^{t_{k+1}}\left(\exp\left(\theta^0\int_0^{t}A(s)\mathrm{d}s\right)-\exp\left(\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)\right.\\
&\quad\left.\left.+\exp\left(\theta\int_0^{\infty}A(s)\mathrm{d}s\right)-\exp\left(\theta\int_0^{t}A(s)\mathrm{d}s\right)\right)\mathrm{d}t\right\}.
\end{aligned} \tag{6.1.23}$$

当 $\theta-\theta^0\geqslant\Delta>0$ 时，

$$\begin{aligned}
&\left(\exp\left(\theta\int_0^{\infty}A(s)\mathrm{d}s\right)-\exp\left(\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)\right)^2\\
&\quad\geqslant\exp\left(2\theta^0\int_0^{\infty}A(s)\mathrm{d}s\right)\left(\exp\left(\varepsilon\int_0^{\infty}A(s)\mathrm{d}s\right)-1\right)^2\\
&\quad\triangleq\delta(\varepsilon)>0,
\end{aligned}$$

注意到当 $k\to\infty$ 时，式(6.1.23)中最后一个从 t_k 到 t_{k+1} 的积分趋于 0，所以式(6.1.12)对一切充分大的 n 成立. 而引理 2.4.1 实际上也只要求式(6.1.12)对一切 $n\geqslant N$ 成立，N 可以任意大. 所以定理 3.1.1 照样可用. 因此，当加权因子 $\{a_k\}$ 满足 A6.1.2，量测噪声满足 A6.1.3 时，式(6.1.5)～(6.1.7)给出的 $\theta_k\xrightarrow[k\to\infty]{}\theta^0$.

上面我们用了离散时间算法来估计 θ^0，我们也可以用连续时间算法.

$$d\theta_t=a_t(dy_t-g_t(x_t(\theta_t))dt), \tag{6.1.24}$$

以及它的变界截尾算法(5.2.7)～(5.2.11).

在§5.2 中的回归函数 $h(\cdot)$ 不依赖于时间，而现在回归函数是时变函数

$$h_t(\theta)=g_t(x_t(\theta^0))-g_t(x_t(\theta)). \tag{6.1.25}$$

为了使引理 5.2.3 成立，我们用 Lyapunov 函数的关键条件(5.2.4)来估计式(5.2.27)右端的第一项. 对连续时间算法的时变回归函数，把条件(5.2.4)相应地改为下面条件.

1) 对任意 $\theta\neq\theta^0$ 及常数 $c>0$，存在充分大的 t' 及充分小的 T，只要 $\|\theta'-\theta\|\leqslant cT$，必有

$$h_s^{\tau}(\theta')v_\theta(\theta)<-a,\ a>0,\ \forall s\in[t,\ s(t,\ T)],\ \forall t\geqslant t', \tag{6.1.26}$$

这里 $v_\theta(\theta)$ 表示 $v(\theta)$ 的梯度，a 可依赖于 θ 及 cT，当 t' 面定后，a 不依赖于 t.

2) 条件 A6.1.1 改成：对任意 $\theta\in\Theta$，方程(6.1.1)对任意初值有大范围解 $x_t(\theta)$，$h_t(\theta)$ 满足局部 Lipschitz 条件

$$\|h_t(\theta_1)-h_t(\theta_2)\|^2 I_{[\|\theta_1\|<n,\|\theta_2\|<n]}\leqslant L_n\|\theta_1-\theta_2\|^2,\ \forall t\geqslant t',$$

L_n 可依赖于 n，但不依赖于 t. 在 θ 的任一有界集 $A\subset\Theta$ 上，$g_t(x_t(\theta))$ 有界.

还要用到下面两个条件.

3) $a_t>0,\ \int_0^\infty a_t\,dt=\infty,\ \int_0^\infty a_t^2\,dt<\infty,$

4) $\varepsilon_t=\int_{t_0}^{t} F(\theta_s)\,\mathrm{d}\omega_s+\int_{t_0}^{t}\eta_s\,\mathrm{d}s$,

这里$(w_t, \mathscr{F}_t)$是 Wiener 过程, $F(\cdot)$ 满足局部 Lipschitz 条件, $(\eta_t, \mathscr{F}_t)$为适应过程, 并且 $\eta_t \xrightarrow[t\to\infty]{} 0$.

利用定理 5.2.1, 我们有以下结果.

定理 6.1.2[6] 设存在 $R^m\to R$ 的二次可微函数$v(\cdot)$; $v(\theta^0)=0$, $v(\theta)\neq 0\ \forall\theta\neq\theta^0$, 存在常数 c_0, 使 $\|\theta^0\|<c_0$, 并且

$$v(\theta^*)<\inf_{\|\theta\|=c_0} v(\theta).$$

设条件 1)~4) 成立, 那么变界截尾的连续时间算法给出的

$$\theta_t\xrightarrow[t\to\infty]{}\theta^0\quad \text{a.s.}.$$

对例 1, 若用连续时间算法, 则

$$h_t(\theta)=\Phi_{t,0}(\theta^0-\theta).$$

当 $\|\Phi_{t,0}\|$ 有界, $\Phi_{t,0}>\delta I$, $\delta>0$, $\forall t\in[0,\infty)$, $g_t(x)\equiv x$ 时, 和例 1 类似地可取 $v(\theta)=\|\theta^0-\theta\|^2$, 并验证式(6.1.26)成立.

对例 2, $h_t(\theta)=\left(\exp\left[\theta^0\int_0^t A(s)\mathrm{d}s\right]-\exp\left[\theta\int_0^t A(s)\mathrm{d}s\right]\right)x(0)$,

在例 2 的条件下, 我们仍可证式(6.1.26)成立.

§6.2 再访 ODE 方法

我们在第 2 章中介绍了用 ODE 方法分析递推算法的收敛性. Ljung 是在研究下面的递推算法时引进 ODE 方法的[65]. 我们先叙述 Ljung 所讨论的问题.

设 l-维参数估计 x_k 递推地给出:

$$x_{k+1}=x_k+a_kQ(k,\ x_k,\ \varphi_{k+1}),\tag{6.2.1}$$

φ_{k+1} 是 m-维系统状态:

$$\varphi_{k+1}=A(x_k)\varphi_k+B(x_k)e_{k+1},\tag{6.2.2}$$

$\{e_{k+1}\}$ 是系统噪声. 初值 φ_0, x_0 任意给出.

记

$$D_s=\{x\text{: } A(x)\quad 稳定\},\tag{6.2.3}$$

设 D_R 是 D_s 的开连通子集.

文献[65]用了三组不同的条件.我们叙述其中的一组. 设

A 1 $\|e_k\|<e<\infty$, a.s.$\forall k$, $\mathrm{E}e^2<\infty$;

A 2 $Q(k, x, \varphi)$ 对 $x\in D_R$ 及 φ 满足局部 Lipschitz 条件, Lipschitz 系数不依赖于 k;

A 3 $A(\cdot)$ 及 $B(\cdot)$ 在 D_R 中 Lipschitz 连续;

A 4 定义

$$\varphi_{k+1}(x)=A(x)\varphi_k(x)+B(x)e_{k+1},\ \varphi_0(x)=0,\ \forall x\in D_R, \tag{6.2.4}$$

对 $x\in D_R$ 存在以下极限

$$\lim_{k\to\infty}\mathrm{E}Q(k,\ x,\ \varphi_{k+1}(x))\triangleq f(x); \tag{6.2.5}$$

A 5 $a_i>0$, $\sum_{i=0}^{\infty}a_i=\infty$;

A 6 对某个 $p>0$,

$$\sum_{i=1}^{\infty}a_i^p<\infty;$$

A 7 $\{e_k\}$ 相互独立;

A 8 a_k 下降;

A 9 $\limsup_{k\to\infty} a_k^{-1}-a_{k-1}^{-1}<\infty$.

文献[65]在证明下述命题时,引进了 ODE 方法.

命题 设 A1～A9 成立,x_k 一直在 D_R 内,无穷多次出现 $\|x_k\|<c<\infty$, $\|\varphi_k\|<c$(c 可依赖样本),并且微分方程

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t)=f(x(t)) \tag{6.2.6}$$

渐近稳定,那么

$$x_k\xrightarrow[k\to\infty]{}J\triangleq\{x: f(x)=0\}\quad \text{a.s.}.$$

这个命题的证明(见文献[65])相当复杂,并且不易读懂.

下面我们讨论较一般的情形. 设 $Q(k,\ x_k,\ \varphi_{k+1})$ 带噪声地观测到:

$$y_{k+1}=Q(k,\ x_k,\ \varphi_{k+1})+\nu_{k+1}. \tag{6.2.7}$$

那么和式(6.2.1)相对应的算法是

$$\begin{aligned}x_{k+1}&=x_k+a_k y_{k+1}=x_k+a_k(Q(k,\ x_k,\ \varphi_{k+1})+\nu_{k+1})\\&=x_k+a_k(f(x_k)+\varepsilon_{k+1}),\end{aligned} \tag{6.2.8}$$

这里

$$e_{k+1}=Q(k,\ x_k,\ \varphi_{k+1})-f(x_k)+\nu_{k+1} \tag{6.2.9}$$

和式(3.1.3)、(3.1.4)类似地把式(6.2.8)变界截尾，得到如下算法

$$\sigma_k=\sum_{i=1}^{k-1} I_{[\|x_i+a_i(f(x_i)+\varepsilon_{i+1})\|>M_{\sigma_i}]},\ \sigma_0=0, \tag{6.2.10}$$

$$\begin{aligned}x_{k+1}=&(x_k+a_k(f(x_k)+\varepsilon_{k+1}))I_{[\|x_k+a_k(f(x_k)+\varepsilon_{k+1})\|\leqslant M_{\sigma_k}]}\\&+x^* I_{[\|x_k+a_k(f(x_k)+e_{k+1})\|>M_{\sigma_k}]}.\end{aligned} \tag{6.2.11}$$

下面的定理[10] 不仅讨论了较一般的量测(6.2.7)，并且把上述命题中的条件 A7、A8、A9 去掉了．值得指出，去掉 A7 是一个实质性改进，现在对 e_k 之间的相关性不加作任何限制．

定理 6.2.1 设条件 A1、A2、A3、A4、A5、A6 成立，把方程(6.2.6)的渐近稳定性具体化为：存在二次连续可微函数 $v(\cdot)$，它在 J 上取值 $v(J)$ 不稠于任何区间，并且只要 $d(x,\ J)>0$，就有 $d(v(x),\ v(J))>0$，同时

$$\sup_{\delta<d(x,J)<\Delta} f^{\tau}(x)v_x(x)<0,\ \forall\delta,\ \Delta;\quad 0<\delta<\Delta. \tag{6.2.12}$$

还设存在常数 c_0，使 $\|x^*\|<c_0$，$v(x^*)<\inf\limits_{\|x\|=c_0} v(x)$．对式(6.2.7)中的量测噪声 $\{\nu_k\}$ 要求

$$\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\sum_{i=k}^{m(k,T)} a_i\nu_{i+1}=0,\quad \text{a.s.}, \tag{6.2.13}$$

这里 $m(k,\ T)$ 由式(2.4.3)定义．除了一个零测集外，对固定的样本，设由式(6.2.9)～(6.2.11)给出的 $\{x_k\}$ 一直在 D_R 内，$\overline{D}_R\subset D_s$，并且对任一收敛子列 $\{x_{n_k}\}$ 相对应的 $\{\varphi_{n_k}\}$ 一致有界：$\|\varphi_{n_k}\|<c$．那么对这个样本(ω)

$$d(x_k,\ J)\xrightarrow[k\to\infty]{}0. \tag{6.2.14}$$

证明 证明的基本思路如下：先去掉一个例外的零测集，在余集上验证式(6.2.9)所表达的 $\{\varepsilon_k\}$ 满足条件 A2.4.3，利用定理 3.1.1 就得出式(6.2.14)。

设 r 为任一具有有理分量的 l-维矢量。对任一 $x\in R^l$，任一 $\varepsilon>0$，总存在 r，使 $\|x-r\|<\varepsilon$。记

$$\omega_{i+1,1}=Q(i,\ x_i,\ \varphi_{i+1})-Q(i,\ x,\ \varphi_{i+1}(x)), \tag{6.2.15}$$

$$\omega_{i+1,2}=Q(i,\ r,\ \varphi_{i+1}(r))-\mathrm{E}Q(i,\ r,\ \varphi_{i+1}(r)), \tag{6.2.16}$$

$$\omega_{i+1,3}=\mathrm{E}Q(i,\ x,\ \varphi_{i+1}(x))-f(x_i), \tag{6.2.17}$$

$$\omega^1_{i+1}=\sum_{j=1}^{3}\omega_{i+1,j},$$

$$\omega^2_{i+1}=Q(i,\ x,\ \varphi_{i+1}(x))-\mathrm{E}Q(i,\ x,\ \varphi_{i+1}(x))-\omega_{i+1,2}, \tag{6.2.18}$$

那么

$$\omega_{i+1}\triangleq Q(i,\ x_i,\ \varphi_{i+1})-f(x_i)=\omega^1_{i+1}+\omega^2_{i+1}, \tag{6.2.19}$$

$$\varepsilon_{i+1}=\omega_{i+1}+\nu_{i+1}. \tag{6.2.20}$$

第一步

固定 r，我们来证

$$\sum_{i=1}^{\infty}a_i\omega_{i+1,2}<\infty\quad \text{a.s.}, \tag{6.2.21}$$

设在 A6 中 $p\leqslant 2$. 那么 $\sum\limits_{i=1}^{\infty}a_i^2<\infty$. 用 $F_k(y_{k+1},\ \cdots,\ y_{k+n})$，表示给定 $e_1,\ \cdots,\ e_k$ 后，$(e_{k+1},\ \cdots,\ e_{k+n})$的条件分布，而用 $F(y_1,\ \cdots,\ y_k)$表示 $e_1,\ \cdots,\ e_k$ 的分布函数. 于是

$$\begin{aligned}&F(y_1,\ \cdots,\ y_{k+n})\\&\quad=F(y_1,\ \cdots,\ y_k)F_k(y_{k+1},\ \cdots,\ y_{k+n}).\end{aligned} \tag{6.2.22}$$

因为 $r\in\overline{D}_R\subset D_s$，所以存在 $\lambda(r)\in(0,\ 1)$，使

$$\|A^n(r)\|\leqslant c_1(r)\lambda^n(r),\quad \forall n\geqslant 0, \tag{6.2.23}$$

注意到

$$\varphi_{k+n}(r)=A^n(r)\varphi_k(r)+\sum_{j=k+1}^{k+n}A^{k+n+j}(r)B(r)e_j,\ \varphi_0(r)=0, \tag{6.2.24}$$

由式(6.2.23)就有

$$
\begin{aligned}
&\|\mathrm{E}(\omega_{k+n,2}|e_1, \cdots, e_k)\| \\
&=\Big\|\underset{y_{k+1}\cdots y_{k+n}}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}} Q\Big(k+n-1, r, \sum_{j=1}^{k} A^{k+n-j}(r)B(r)e_j \\
&\quad+\sum_{j=k+1}^{k+n} A^{k+n-j}(r)B(r)y_j\Big)\mathrm{d}F_k(y_{k+1}, \cdots, y_{k+n}) \\
&\quad-\underset{y_1\cdots y_{k+n}}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}} Q\Big(k+n-1, r, \sum_{j=1}^{k+n} A^{k+n-j}(r)B(r)y_j\Big) \\
&\quad\cdot\mathrm{d}F(y_1, \cdots, y_k)\mathrm{d}F_k(y_{k+1}, \cdots, y_{k+n})\Big\| \\
&\leqslant c_2(r)\lambda^n(r)\underset{y_1\cdots y_k}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}\sum_{j=1}^{k}\lambda^{k-j}(r)\|e_j-y_j\| \\
&\quad\cdot\mathrm{d}F(y_1, \cdots, y_k) \\
&\leqslant c_2(r)\lambda^n(r)\sum_{j=1}^{k}\lambda^{k-j}(r)(\|e_j\|+\mathrm{E}\|e_j\|) \\
&\leqslant c_3(r)\lambda^n(r),
\end{aligned}
\tag{6.2.25}
$$

这里$C_2(r)<\infty$, $\mathrm{E}C_3^2(r)<\infty$, 在式(6.2.5)的推导中，第一个不等式用了条件A 2、$\{e_k\}$ 和 $\{\varphi_k(x)\}$ 的有界性，而第三个不等式用了条件A 1.

这样，我们就有

$$
\begin{aligned}
&\|E(\omega_{k,1}\omega_{k+n,2}^{\tau})\| \\
&=\|\mathrm{E}[\omega_{k,2}\mathrm{E}(\omega_{k+n,2}^{\tau}|e_1, \cdots, e_k)]\|\leqslant c_3(r)\lambda^n(r)\mathrm{E}\|\omega_{k,2}\| \\
&=c_3(r)\lambda^n(r)\underset{y_1\cdots y_k}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}\Big\|Q\Big(k, r, \sum_{j=1}^{k} A^{k-j}(r)B(r)y_j\Big) \\
&\quad-\underset{z_1\cdots z_k}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}} Q\Big(k, r, \sum_{j=1}^{k} A^{k-j}(r)B(r)z_j\Big)\mathrm{d}F(z_1, \cdots, z_k)\Big\| \\
&\quad\cdot\mathrm{d}F(y_1, \cdots, y_k) \\
&\leqslant c_4(r)\lambda^n(r)\underset{y_1\cdots y_k\ z_1\cdots z_k}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}} \\
&\quad\cdot\Big\|\sum_{j=1}^{k} A^{k-j}(r)B(r)(y_j-z_j)\Big\|
\end{aligned}
$$

$$\cdot \mathrm{d}F(z_1, \cdots, z_k)\mathrm{d}F(y_1, \cdots, y_k)$$

$$\leqslant c_1(r)c_4(r)\lambda^n(r)\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}$$

$$\cdot \sum_{j=1}^{k}\lambda^{k-j}(r)\|B(r)\|(\|y_j\|+\|z_j\|)$$

$$\cdot \mathrm{d}F(z_1, \cdots, z_k)\mathrm{d}F(y_1, \cdots, y_k)$$

$$\leqslant c_5(r)\lambda^n(r), \tag{6.2.26}$$

最后的不等式是因为 $\mathrm{E}e^2<\infty$. $c_i(r)$, $i=1, 2, \cdots$ 都不依赖于时间.

由式(6.2.26)便得

$$\mathrm{E}\left\|\sum_{i=1}^{\infty}a_i\omega_{i+1,2}\right\|^2$$

$$=\sum_{i=1}^{\infty}a_i^2\mathrm{E}\|\omega_{i+1,2}\|^2+2\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}a_ia_{i+k}\mathrm{E}\omega_{i+1,2}^{\tau}\omega_{i+k+1,2}$$

$$\leqslant C_5(r)\sum_{i=1}^{\infty}a_i^2+C_5(r)\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}(a_i^2+a_{i+k}^2)\lambda^k(r)<\infty, \tag{6.2.27}$$

由此便得式(6.2.21).

当 A6 中 $p>2$ 时, 取 $q=\min\{$偶整数 k: $k\geqslant p\}$. 显然 $\sum_{i=1}^{\infty}a_i^q<\infty$. 和式(6.2.22)~(6.2.27)类似地可证

$$\mathrm{E}\left\|\sum_{i=1}^{\infty}a_i\omega_{i+1,2}\right\|^q<\infty,$$

由此得式(6.2.21).

对每个 r, $S_r\triangleq\left(\omega: \sum_{i=1}^{\infty}a_i\omega_{i+1,2}<\infty\right)$, $PS_r=1$. 由于 r 为可列个, 所以 $P\bigcap_r S_r=1$. 用 S_v 表示

$$S_v=\left(\omega:\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\sum_{i=k}^{m(k,T)}a_iv_{i+1}=0\right),$$

那么

$$P(S_v\bigcap_r S_r)=1.$$

下面把零概率事件 $\Omega\backslash(S_v\bigcap_r S_r)$去掉, 总认为 $\omega\in(S_v\bigcap_r S_r)$.

固定 ω，设 $x_{n_k} \xrightarrow[k\to\infty]{} x$.

第二步

用 $K(\cdot)$ 表示局部 Lipschitz 常数:

$$(\|Q(k, x_1, \varphi_1)-Q(k, x_2, \varphi_2)\|+\|A(x_1)-A(x_2)\| \\ +\|B(x_1)-B(x_2)\|)I_{[\|x_1\|\leqslant N,\|x_2\|\leqslant N,\|\varphi_1\|\leqslant N,\|\varphi_2\|\leqslant N]} \\ \leqslant K(N)(\|x_1-x_2\|+\|\varphi_1-\varphi_2\|). \quad (6.2.28)$$

记 $$\Delta_m=\max_{n_k\leqslant i\leqslant m}\|x_i-x\|,\ \Delta_{n_k-1}=0.$$

我们来证存在 $c_6(x)$ 使

$$\|\varphi_i\|\leqslant c_6(x)+c_6(x) \\ \times\sum_{s=n_k}^{i-1}(\lambda(x)+c_1(x)\Delta_{i-1}K(\Delta_{i-1}+\|x\|))^{i-s}, \\ i\geqslant n_k+1. \quad (6.2.29)$$

把式(6.2.2)写成

$$\varphi_i=A(x)\varphi_{i-1}+(A(x_{i-1})-A(x))\varphi_{i-1}+B(x_{i-1})e_i. \quad (6.2.30)$$

注意到 $\|\varphi_{n_k}\|\leqslant c$，由式(6.2.23)及式(6.2.28)得

$$\|\varphi_i\|\leqslant c_1(x)\lambda^{i-n_k}(x)\|\varphi_{n_k}\| \\ +c_1(x)\sum_{j=n_k+1}^{i}\lambda^{i-j}(x)\{\|A(x_{j-1})-A(x)\|\cdot\|\varphi_{j-1}\| \\ +\|B(x_{j-1})-B(x)\|+\|B(x)\|\}\|e_j\| \\ \leqslant c_1(x)C\lambda^{i-n_k}+c_1(x)K(\Delta_{i-1}+\|x\|) \\ \cdot\sum_{j=n_k+1}^{i}\lambda^{i-j}(x)\Delta_{j-1}(\|\varphi_{j-1}\|+1) \\ +c_1(x)\sum_{j=n_k+1}^{i}\lambda^{i-j}(x)\|B(x)\|e, \quad i\geqslant n_k.$$

所以

$$1+\|\varphi_i\|\leqslant c_6(x)+c_1(x)K(\Delta_{i-1}+\|x\|) \\ \cdot\sum_{j=n_k+1}^{i}\lambda^{i-j}(x)\Delta_{j-1}(1+\|\varphi_{j-1}\|).$$

记 $h_i=\lambda^{-i}(x)(1+\|\varphi_i\|)$，那么从上式得

$$h_i \leqslant c_6(x)\lambda^{-i}(x) + \frac{c_1(x)K(\Delta_{i-1}+\|x\|)}{\lambda(x)} \sum_{j=n_k+1}^{i} \Delta_{j-1}h_{j-1}. \tag{6.2.31}$$

由定理 1.3.5 知

$$h_i \leqslant c_6(x)\lambda^{-i}(x) + \frac{c_1(x)K(\Delta_{i-1}+\|x\|)\Delta_{i-1}}{\lambda(x)} \cdot c_6(x) \sum_{s=n_k}^{i-1}\left(1+\frac{c_1(x)K(\Delta_{i-1}+\|x\|)\Delta_{i-1}}{\lambda(x)}\right)^{i-s-1}\lambda^{-s}(x),$$

所以

$$\begin{aligned}1+\|\varphi_i\| &\leqslant c_6(x)+c_1(x)c_6(x)K(\Delta_{i-1}+\|x\|)\Delta_{i-1} \\ &\quad\times \sum_{s=n_k}^{i-1}(\lambda(x)+c_1(x)K(\Delta_{i-1}+\|x\|)\Delta_{i-1})^{i-s-1} \\ &\leqslant c_6(x)+c_6(x) \\ &\quad\times \sum_{s=n_k}^{i-1}(\lambda(x)+c_1(x)\Delta_{i-1}K(\Delta_{i-1}+\|x\|))^{i-s},\end{aligned}$$

由此得式(6.2.29).

第三步

记 $c(x)=1+\|f(x)\|$，取 $T(x)\in(0, 1]$ 充分小，使

$$\lambda_1(x) \triangleq (\lambda(x)+c_1(x)K(c(x)T(x)+\|x\|))c(x)T(x) < 1. \tag{6.2.32}$$

我们现在来证，如果对某个 m: $n_k \leqslant m \leqslant m(n_k, T(x))$，

$$\Delta_m \leqslant c(x)T(x), \tag{6.2.33}$$

那么存在 $c_7(x)$ 和 $c_8(x)$ 使

$$\left\|\sum_{i=n_k}^{m} a_i\omega_{i+1,1}\right\| \leqslant \max_{n_k\leqslant i\leqslant m} a_i c_7(x)\frac{\lambda(x)}{1-\lambda(x)} + c_8(x)T^2(x). \tag{6.2.34}$$

由式(6.2.29)、(6.2.32)及(6.2.33)知

$$\|\varphi_{i+1}\| \leqslant \frac{c_6(x)}{1-\lambda_1(x)}, \quad \forall i:\quad n_k \leqslant i \leqslant m. \tag{6.2.35}$$

由式(6.2.23)、(6.2.24)及 $\|e_j\|\leqslant e$, $\forall j\geqslant 1$ 知

$$\|\varphi_i(x)\| \leqslant c_9(x), \quad \forall i\geqslant 1. \tag{6.2.36}$$

根据式(6.2.35)、(6.2.36)及(6.2.33)，记 $\tilde{\varphi}_i=\varphi_i-\varphi_i(x)$，对

$\omega_{i+1,1}$ 用条件 A 2, 就有

$$\|\omega_{i+1,1}\| \leqslant K\left(\Delta_i + \|x\| + \frac{c_6(x)}{1-\lambda_1(x)} + c_9(x)\right)(\Delta_i + \tilde{\varphi}_{i+1})$$
$$\leqslant K\left(c(x)T(x) + \|x\| + \frac{c_8(x)}{1-\lambda_1(x)} + C_9(x)\right)$$
$$\cdot (c(x)T(x) + \tilde{\varphi}_{i+1}), \quad n_k \leqslant i \leqslant m. \tag{6.2.37}$$

显然对 i: $n_k \leqslant i \leqslant m$, 有

$$\tilde{\varphi}_{i+1} = A(x)\tilde{\varphi}_i + [A(x_i) - A(x)]\varphi_i$$
$$+ [B(x_i) - B(x)]e_{i+1}. \tag{6.2.38}$$

注意到 $\|\tilde{\varphi}_{n_k}\| \leqslant c + c_0(x) \triangleq c_{10}(x)$, 由式(6.2.23)、(6.2.33)和(6.2.35), 我们得到

$$\|\tilde{\varphi}_{i+1}\| \leqslant c_1(x)\, c_{10}(x) \lambda^{i+1-n_k}(x) + c_1(x) \sum_{j=n_k+1}^{i+1} \lambda^{i+1-j}(x)$$
$$\times [\|A(x_{j-1}) - A(x)\| \cdot \|\varphi_{j-1}\| + \|B(x_{j-1}) - B(x)\| e]$$
$$\leqslant c_1(x) c_{10}(x) \lambda^{i+1-n_k}(x) + c_1(x) K(\Delta_i + \|x\|)$$
$$\times \sum_{j=n_k+1}^{i+1} \lambda^{i+1-j}(x) \Delta_{j-1}(\|\varphi_{j-1}\| + e)$$
$$\leqslant c_1(x) c_{10}(x) \lambda^{i+1-n_k}(x) + c_1(x) K(c(x)T(x) + \|x\|)$$
$$\times\ c(x)T(x)\left(\frac{c_6(x)}{1-\lambda_1(x)} + e\right)\frac{1}{1-\lambda(x)}. \tag{6.2.39}$$

由式(6.2.37)、(6.2.39)知, 存在 $C_7(x)$, $C_8(x)$ 使

$$\|\omega_{i+1,1}\| \leqslant c_7(x) \lambda^{i+1-n_k}(x) + c_8(x) T(x), \quad n_k \leqslant i \leqslant m.$$

由此便得式(6.2.34), 所以如果式(6.2.33)对任意 m: $n_k \leqslant m \leqslant m(n_k, T(x))$ 成立, 那么对 $j=1$ 成立

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k,T)} a_i \omega_{i+1,j} \right\| = 0, \quad \forall T_k \in [0, T]. \tag{6.2.40}$$

第四步

由此便知, 为证式(6.2.40)对 $j=1$ 成立, 我们只要证可选 $T(x)$ 充分小, 使对充分大的 k_1 有

$$\Delta_{m(n_k, T(x))} \leqslant C(x) T(x), \quad \forall k \geqslant k_1. \tag{6.2.41}$$

我们用归纳法来证这一点. 由于 $x_{n_k}\xrightarrow[k\to\infty]{}x$,所以对任意固定的 $T(x)$,只要 k 充分大就有

$$\Delta_{n_k}\leqslant\frac{1}{2}T(x)\quad(<c(x)T(x)).\tag{6.2.42}$$

设我们已证

$$\Delta_j\leqslant c(x)T(x),\quad\forall j:\quad n_k\leqslant j\leqslant m<m(n_k,\ T(x)).\tag{6.2.43}$$

记

$$\begin{aligned}g_k\triangleq&\max_{n_k\leqslant i\leqslant j}a_jc_7(x)\frac{\lambda(x)}{1-\lambda(x)}+\left\|\sum_{i=n_k}^{j}a_i\omega_{i+1,2}\right\|+\left\|\sum_{i=n_k}^{j}a_i\nu_{i+1}\right\|\\&+\max_{n_k\leqslant i\leqslant j}\|\mathrm{E}Q(i,\ x,\ \varphi_{i+1}(x))-f(x)\|T(x)\\&+\max_{n_k\leqslant i\leqslant j}\|\omega_{i+1}^2\|T(x).\end{aligned}$$

由式(6.2.34)我们有

$$\begin{aligned}&\left\|\sum_{i=n_k}^{j}a_i(Q(i,\ x_i,\ \varphi_{i+1})+\nu_{i+1})\right\|\\&\leqslant\left\|\sum_{i=n_k}^{j}a_i(\omega_{i+1,1}+\omega_{i+1,2})\right\|+\left\|\sum_{i=n_k}^{j}a_i\nu_{i+1}\right\|\\&\quad+\left\|\sum_{i=n_k}^{j}a_i(\mathrm{E}(Q(i,\ x,\ \varphi_{i+1}(x))-f(x))\right\|\\&\quad+\sum_{i=n_k}^{j}a_i\|f(x)\|+\left\|\sum_{i=n_k}^{j}a_i\omega_{i+1}^2\right\|\\&\leqslant g_k+(c_8(x)T(x)+\|f(x)\|)T(x).\end{aligned}\tag{6.2.44}$$

取 $T(x)$ 充分小, 使得式(6.2.32)成立, 并使 $C_8(x)T(x)<\frac{1}{4}$. 固定这个 $T(x)$, 取 r 充分接近 x, 那么由式(6.2.21)、式(6.2.13)、条件 A4, 并注意到 $a_i\xrightarrow[i\to\infty]{}0$, 便知可取不依赖 j 的充分大 k_1, 使

$$g_k/T(x)<\frac{1}{4},\quad\forall k\geqslant k_1.$$

这样由式(6.2.44)就有

$$\left\|\sum_{i=n_k}^{j}a_i(Q(i,\ x_i,\ \varphi_{i+1})+\nu_{i+1})\right\|$$

$$\leqslant\left(\frac{1}{2}+T\|f(x)\|\right)T(x),\ \forall k\geqslant k_1,\ \forall j:\ n_k\leqslant j\leqslant m.$$

(6.2.45)

不失一般性可认为

$$1+\|f(x)\|+\|x\|<M\sigma_{n_k},\quad \forall k\geqslant k_1. \tag{6.2.46}$$

那么由式(6.2.42)、(6.2.45)、(6.2.46)便知对 $\forall j:n_k\leqslant j\leqslant m$

$$\left\|x_{n_k}+\sum_{i=n_k}^{j}a_i(Q(i,\ x_i,\ \varphi_{i+1})+\nu_{i+1})\right\|$$
$$<(1+\|f(x)\|)T(x)+\|x\|<M\sigma_{n_k}.$$

所以在 j 时刻，$n_k+1\leqslant j\leqslant m+1$ 没有截尾，因此

$$x_{m+1}=x_m+a_m(Q(m,\ x_m,\ \varphi_{m+1})+\nu_{m+1}).$$

由式(6.2.42)、(6.2.45)知

$$\|x_{m+1}-x\|\leqslant\|x_{n_k}-x\|$$
$$+\left\|\sum_{i=n_k}^{m}a_i(Q(i,\ x_i,\ \varphi_{i+1})+\nu_{i+1})\right\|$$
$$<c(x)T(x),$$

即 $\Delta_{m+1}<c(x)T(x)$. 这样我们就用归纳法证明了式(6.2.41). 所以式(6.2.40)对 $j=1$ 成立，从式(6.2.21)知式(6.2.40)在所考察的 ω 上对 $j=2$ 也成立. 我们来证它对 $j=3$ 也成立.

第五步

注意到对 x、$y\in D_R$, $\overline{D}_R\subset D_3$，成立式(6.2.23)，所以

$$\lim_{x\to y}\lim_{n\to\infty}\sum_{j=1}^{n}\|A^j(x)-A^j(y)\|$$
$$=\lim_{x\to y}\lim_{n\to\infty}\left[\sum_{j=1}^{N}\|A^j(x)-A^j(y)\|\right.$$
$$\left.+\sum_{j=N+1}^{n}(\|A^j(x)\|+\|A^j(y)\|)\right]$$
$$=\lim_{x\to y}\sum_{j=N+1}^{n}(\|A^j(x)\|+\|A^j(y)\|)\xrightarrow[N\to\infty]{}0.$$

因此由条件 A1～A4 知，$f(\cdot)$在 D_R 内为连续函数.

由条件 A4 及式(6.2.41)知对任一 $T_k\in[0,\ T(x)]$，有

$$\limsup_{k\to\infty}\frac{1}{T(x)}\left\|\sum_{i=n_k}^{m(n_k,T_k)}a_i\omega_{i+1,3}\right\|$$

$$\leqslant \limsup_{k\to\infty} \frac{1}{T(x)} \left\{ \left\| \sum_{i=n_k}^{m(n_k,T_k)} a_i(f(x)-f(x_i)) \right\| + \left\| \sum_{i=n_k}^{m(n_k,T_k)} a_i[\mathrm{E}Q(i,\ x,\ \varphi_{i+1}(x))-f(x)] \right\| \right\}$$

$$=\limsup_{k\to\infty} \frac{1}{T(x)} \left\| \sum_{i=n_k}^{m(n_k,T_k)} a_i(f(x)-f(x_i)) \right\|$$

$$\leqslant \limsup_{k\to\infty} \max_{n_k\leqslant i\leqslant m(n_k,T(x))} \|f(x)-f(x_i)\|$$

$$\leqslant \max_{\substack{y\in D_R\\ \|x-y\|\leqslant c(x)T(x)}} \|f(x)-f(y)\|.$$

由于 $f(\cdot)$在 D_R 内连续，所以上式当 $T(x)\to 0$ 时，趋于 0．所以式(6.2.40)对 $j=3$ 也成立．

注意到对固定的 x，可选 r 使 ω_{i+1}^2 任意小，所以 $\{e_k\}$ 满足 A 2.4.3，利用定理 3.1.1，就得式(6.2.14)． □

上面用随机逼近收敛性定理，证明了算法收敛到 $f(\cdot)$ 的零点．下面我们证明当 x_k 收敛时，那么极限就是 $f(\cdot)$的零点．

定理 6.2.2[10] 设条件 A 1～A 6 成立，A1 中的 e 为常数，式(6.2.13)的收敛对 ω 一致．如果存在 $x^0\in D_R$，使 $PA_\rho>0$，$\forall\rho>0$，这里 $A_\rho\triangleq\{\omega: x_k\to\mathscr{B}(x^0,\ \rho)\}$，$\mathscr{B}(x^0,\ \rho)$是以 x^0 为中心，ρ 为半径的球，那么 $f(x^0)=0$．

证明 从式(6.2.21)知

$$\sum_{i=1}^{\infty} a_i(Q(i,\ x^0,\ \varphi_{i+1}(x^0))-\mathrm{E}Q(i,\ x^0,\ \varphi_{i+1}(x^0))) <\infty \quad \text{a.s.,} \tag{6.2.47}$$

用 Ω' 表示 ω 集，在 Ω' 上，式(6.2.13)对 $T=\frac{1}{j}$，$j=1,\ 2,\ 3,\ \cdots$ 都成立，并且式(6.2.47)也成立，显然 $P\Omega'=1$．

由于 D_R 为开集，$D_R\subset D_s$，那么对充分小的 ρ_0，存在 $\mu\in(0,1)$及常数 C_ρ，使

$$\left\| \prod_{i=1}^{m} A(y_i) \right\| \leqslant c_{\rho_0}\mu^m,\ \forall y_i\in\mathscr{B}(x^0,\ \rho_0). \tag{6.2.48}$$

设 $\rho_1\in(0,\ \rho_0)$．对 $\omega\in A_{\rho_1}$，只要 k_1 充分大，对 $k\geqslant k_1$，$x_k\in A_{\rho_0}$．由于 $\|e_i\|$ 小于一常数，所以存在常数 c 使得

$$\|\varphi_k\|<c,\ \|\varphi_k(x^0)\|<c,\ \forall\omega\in A_{\rho_1}. \tag{6.2.49}$$

考察 $\omega\in A'_{\rho_1}\triangleq A_{\rho_1}\cap\Omega'$. 显然对 $\omega\in A'_{\rho_1}$, 截尾次数有穷, 对 $\omega\in A'_{\rho_1}$, 由条件 A 4式(6.2.13)及(6.2.47)知

$$\begin{aligned}&\limsup_{k\to\infty} j\|x_{m(k,\frac{1}{j})+1}-x_k\|\\&=\limsup_{k\to\infty}\left\|j\sum_{i=k}^{m\left(k,\frac{1}{j}\right)}a_i(Q(i,\ x_i,\ \varphi_{i+1})\right.\\&\qquad\left.-Q(i,\ x^0,\ \varphi_{i+1}(x^0))+f(x^0))\right\|.\end{aligned} \tag{6.2.50}$$

由于式(6.2.13)是对 ω 一致收敛, 由式(6.2.49)知存在 c_1, 对充分大的 k 有

$$\begin{aligned}\|x_{m+1}-x_k\|&=\left\|\sum_{i=k}^{m}a_i(Q(i,\ x_i,\ \varphi_{i+1})+\nu_{i+1})\right\|\\&\leqslant\frac{c_1}{j},\quad\forall m: k\leqslant m\leqslant m\left(k,\ \frac{1}{j}\right),\ \forall\omega\in A'_{\rho_1}.\end{aligned} \tag{6.2.51}$$

记 $\tilde{\varphi}_{i+1}=\varphi_{i+1}-\varphi_{i+1}(x^0)$. 由式(6.2.48)、(6.2.51)并注意到式(6.2.38)就有常数 c_2, c_3, 使对一切 $\omega\in A'_{\rho_1}$ 有

$$\|\tilde{\varphi}_{i+1}\|\leqslant c_2\mu^{i+1-k}+\frac{c_3}{j},\ \forall i: k\leqslant i\leqslant m\left(k,\ \frac{1}{j}\right). \tag{6.2.52}$$

由式(6.2.49)、(6.2.51)、(6.2.52)及条件 A 2 就知存在常数 c_4、c_5 使对一切 $\omega\in A'_{\rho_1}$ 有

$$\begin{aligned}&\|Q(i,\ x_i,\ \varphi_{i+1})-Q(i,\ x^0,\ \varphi_{i+1}(x^0))\|\\&\quad\leqslant c_4\mu^{i+1-k}+\frac{c_5}{j},\ \forall i: k\leqslant i\leqslant m\left(k,\ \frac{1}{j}\right).\end{aligned}$$

由式(6.2.50)及上式知

$$\begin{aligned}&\limsup_{k\to\infty} j\|x_{m\left(k,\frac{1}{j}\right)+1}-x_k\|\\&\quad\geqslant-\frac{c_5}{j}+\|f(x^0)\|,\ \forall\omega\in A'_{\rho_1}.\end{aligned} \tag{6.2.53}$$

反设 $\|f(x^0)\|\neq0$. 不妨可认为 ρ_0 已充分小, 使

$$\|f(x)\|>\delta>0,\quad\forall x: \|x-x^0\|\leqslant\rho_0.$$

取 j 充分大使 $\frac{c_5}{j}<\frac{\delta}{4}$，固定这个 j，取 $\rho\leqslant\rho_1$ 充分小使 $2\rho j<\frac{\delta}{4}$。那么对 $\omega\in A'_{\frac{\rho}{2}}\triangleq A_{\frac{\rho}{2}}\cap\Omega'$ 就有

$$\limsup_{k\to\infty} j\|x_{m\left(k,\frac{1}{j}\right)+1}-x_k\|\leqslant 2\rho j<\frac{\delta}{4}.$$

因此对 $\omega\in A'_{\frac{\rho}{2}}$，式(6.2.53)就成为

$$\frac{\delta}{4}>\limsup_{k\to\infty} j\|x_{m\left(k,\frac{1}{j}\right)+1}-x_k\|\geqslant-\frac{\delta}{4}+\delta=\frac{3\delta}{4}.$$

这个矛盾不等式说明反假设不成立．所以 $f(x^0)=0$． □

§6.3 一个适应控制问题

我们把上一节中得到的收敛定理来解决一个适应控制问题．

设系统的理想运转状态为

$$\begin{gathered}\varphi_{k+1}(x^0)=A(x^0)\varphi_k(x^0)+B(x^0)u_k^0+C(x^0)\varepsilon_{k+1},\\ \varphi_k(x^0)\in R^m,\end{gathered}\tag{6.3.1}$$

u_k^0 是控制，ε_{k+1} 是系统噪声，ε_k 对 $\mathscr{F}_k$ 可测，$\mathscr{F}_k\uparrow$．描述系统理想运转状态的参数 x^0 是函数 $f(\cdot)$ 的根，即 $f(x^0)=0$．但对 x^0 的信息只能通过量测

$$y_{k+1}=Q(k,\ x_k,\ \varphi_{k+1})+\nu_{k+1}\tag{6.3.2}$$

来获得，ν_{k+1} 是量测噪声，x_k 是对 x^0 在 k 时刻的估计．由于 x^0 不知道，所以实际的系统为

$$\varphi_{k+1}=A(x_k)\varphi_k+B(x_k)u_k+C(x_k)\varepsilon_{k+1},\tag{6.3.3}$$

而量测量和 $f(\cdot)$ 的联系如下：

$$\mathrm{E}Q(k,\ x,\ \varphi_{k+1}(x))\xrightarrow[k\to\infty]{}f(x),\tag{6.3.4}$$

$\varphi_{k+1}(x)$ 的定义见式(6.3.15)．

设 $Q_1\geqslant0$，$Q_2>0$，φ_k^* 为已知的有界确定性讯号．记

$$\begin{aligned}J(u^0)=\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}[(\varphi_k(x^0)-\varphi_k^*)^{\tau}Q_1(\varphi_k(x^0)\\ -\varphi_k^*)+u_k^{0\tau}Q_2u_k^0],\end{aligned}\tag{6.3.5}$$

$$U^0=\Big\{u^0:\ \sum_{i=0}^{n-1}(\|u_i^0\|^2+\|\varphi_i(x^0)\|^2)=O(n),\ \|\varphi_n(x^0)\|^2$$
$$=o(n)\quad \text{a.s.},\quad u_i^0\in\mathscr{F}_i,\ \forall i\geqslant 0\Big\}\tag{6.3.6}$$

要求在式(6.3.3)中给出 x_k 及 u_k 使

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}[(\varphi_k-\varphi_k^*)^{\tau}Q_1(\varphi_k-\varphi_k^*)+u_k^{\tau}Q_2u_k]$$
$$=\min_{u^0\in U^0}J(u^0).\tag{6.3.7}$$

设 $D^{\tau}D=Q_1$, $(A(x),\ B(x),\ D)$ 能控及能观，那么可从下面的 Riccati 方程定出 $S(x)$:[15]

$$S(x)=A^{\tau}(x)S(x)A(x)-A^{\tau}(x)S(x)B(x)(Q_2+B^{\tau}(x)$$
$$\cdot S(x)B(x))^{-1}B^{\tau}(x)S(x)A(x)+Q_1.\tag{6.3.8}$$

设

$$L(x)=-(Q_2+B^{\tau}(x)S(x)B(x))^{-1}B^{\tau}(x)S(x)A(x),\tag{6.3.9}$$

$$F(x)=A(x)+B(x)L(x),\tag{6.3.10}$$

$$b_k(x)=-\sum_{j=0}^{n}(F^{\tau}(x))^jQ_1\varphi_{k+j}^*,\tag{6.3.11}$$

$$d_k(x)=-(Q_2+B^{\tau}(x)S(x)B(x))^{-1}B^{\tau}(x)b_{k+1}(x).\tag{6.3.12}$$

那么从文献[15]的定理 3.5 知

$$u_k^0\triangleq L(x^0)\varphi_k(x^0)+d_k(x^0)\in U^0,\tag{6.3.13}$$

把这个 u_k^0 用于式(6.3.1)后就导致以下结果：若$(\varepsilon_i,\ \mathscr{F}_i)$为鞅差列，并且

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\varepsilon_i^{\tau}\xrightarrow[n\to\infty]{}R<\infty\quad \text{a.s.},$$

那么

$$\min_{u^0\in U^0}J(u^0)=\operatorname{tr}S(x^0)C(x^0)RC^{\tau}(x^0)$$
$$+\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}[\varphi_i^{*\tau}Q_1\varphi_i^*-b_{i+1}^{\tau}(x^0)B(x^0)(Q_2$$
$$+B^{\tau}(x^0)S(x^0)B(x^0))^{-1}B^{\tau}(x^0)b_{i+1}(x^0)]\tag{6.3.14}$$

我们要用到下面条件.

H1 设 $D^\tau D=Q_1$, M 为一闭有界集, $M\subset R^l$. $(A(x),\ B(x),\ D)$能控能观, $\forall x\in M$, 且 $x^0\in M$.

H2 $\{\varepsilon_i,\ \mathscr{F}_i\}$ 为鞅差列, $\|\varepsilon_i\|<\varepsilon<\infty$ a.s., $\mathrm{E}\varepsilon<\infty$.

H3 对 φ 及 $x\in M$, $Q(k,\ x,\ \varphi)$满足局部 Lipschitz 条件, 并且 Lipschitz 系数不依赖于 k.

H4 对 $x\in M$, $F(x)$, $B(x)$及 $C(x)$满足局部 Lipschitz 条件.

H5 对 $x\in M$, $\lim\limits_{k\to\infty}\mathrm{E}[Q(k,\ x,\ \varphi_{k+1}(x))]=f(x)$, $f(x^0)=0$, 这里

$$\varphi_{k+1}(x)=F(x)\varphi_k(x)+B(x)d_k(x)+C(x)\varepsilon_{k+1},\ \varphi_0(x)=0. \tag{6.3.15}$$

H6 存在二次可微函数 $v(\cdot)$: $R^l\to B$, 使

$$\sup_{\delta<|x-x^0|,x\in M} f^\tau(x)v_x(x)<0,\quad \forall\delta>0, \tag{6.3.16}$$

并对某个 C_0, $\{x:\ \|x\|\leqslant C_0\}\subset\mathring{M}$, ($\mathring{M}$ 表示 M 的内部, $\mathring{M}=(\overline{M^C})^c$),

$$v(x^*)<\inf_{|x|=c_0} v(x),\ x^*\in\{x:\ \|x\|<c_0\}. \tag{6.3.17}$$

H7 量测噪声 ν_k 和 $\mathscr{F}_k$ 适应, 并且当 x_{n_k} 收敛时

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{i=n_k}^{m(n_k,T)} a_i\nu_{i+1}\right\|=0\quad \text{a.s.}.$$

H8 $a_i>0$, $\sum\limits_{i=1}^{\infty}a_i=\infty$, 并对某个 $p>0$, $\sum\limits_{i=1}^{\infty}a_i^p<\infty$.

对 x^0 的估计由下面递推公式给出

$$x_{k+1}=(x_k+a_ky_{k+1})I_{[(x_k+a_ky_{k+1})\in\mathring{M}]}+x^*I_{[(x_k+a_ky_{k+1})\overline{\in}\mathring{M}]}, \tag{6.3.18}$$

定义适应控制:

$$u_k=L(x_k)\varphi_k+d_k(x_k). \tag{6.3.19}$$

定理 6.3.1 设 H1~H8 成立. 把适应控制(6.3.19)用到系统(6.3.3)后, 除了一个零测集外, 如果当子列 x_{n_k} 收敛时, 相应的

子列 φ_{n_k} 有界：$\|\varphi_{n_k}\|<C$，那么 $x_k \xrightarrow[k\to\infty]{} x^0$，$\varphi_k-\varphi_k(x^0)\xrightarrow[k\to\infty]{}0$，并且成立式(6.3.7)，也就是说，对这个样本当 $k\to\infty$ 时，实际系统趋向于理想的运转状态，并且指标也渐近地达到极小值.

证明 由于对 $x\in M$，$(A(x), B(x), D)$ 能控能观，那么 $F(x)$ 稳定，所以存在 $\lambda\in(0, 1)$ 及 $c_1>0$，使

$$\|F^n(x)\|\leqslant c\lambda^n,\ \forall n\geqslant 1,\ \forall x\in M. \qquad (6.3.20)$$

这和 $\{\varphi_k^*\}$ 的有界性一起，便知 $b_k(x)$ 及 $d_k(x)$ 对 k 及 x 有界：$0\leqslant k<\infty$，$x\in M$．所以 $d_k(x_k)$ 对 k 有界.

用(6.3.19)后的实际系统为

$$\varphi_{k+1}=F(x_k)\varphi_k+G(x_k)e_{k+1}, \qquad (6.3.21)$$

这里

$$G(x)=[B(x)\ \vdots\ C(x)],\ e_{k+1}=\begin{bmatrix} d_k(x_k) \\ \varepsilon_{k+1}\end{bmatrix}. \qquad (6.3.22)$$

在 § 6.2 中，把 $A(x)$、$B(x)$ 相应地换成 $F(x)$ 及 $G(x)$ 后，便知 A 1～A 6 成立．那么和定理 6.2.1 一样地证明，在 $S_\nu\bigcap_r S_r$ 上，式(6.2.40)对 $j=1, 2, 3$ 成立．为了证明 $x_k\to x^0$ 就要用定理 3.1.1．注意到在该定理中算法是变界截尾的，但式(6.3.18)是具有固定截尾界的算法．但这里额外地假定了 $x^0\in M$，相应地在 H6 中要求了 c_0 满足 $\{x:\ \|x\|\leqslant c_0\}\subset \mathring{M}$．所以定理 3.1.1 及有关的引理都照样成立．所以在集合

$$S\triangleq\{\omega:\ \text{如果 } x_{n_k} \text{ 收敛，则 } \varphi_{n_k} \text{ 有界}\}$$

上，$x_k\xrightarrow[k\to\infty]{}x^0$．由此便知 $S(x_k)\to S(x^0)$，$F(x_k)\to F(x^0)$，$\|d_k(x_k)-d_k(x^0)\|\xrightarrow[k\to\infty]{}0$，并且存在 f 及 $\mu\in(0, 1)$ 使

$$\|F(x_n)\cdots F(x_k)\|\leqslant f\mu^{n-k},\ \forall k,\ \forall n\geqslant k. \qquad (6.3.23)$$

由此知 $\{\varphi_k\}$ 对 k 有界．记 $\tilde{\varphi}_k\triangleq\varphi_k-\varphi_k(x^0)$，注意到

$$\begin{aligned}\tilde{\varphi}_{k+1}=&F(x^0)\tilde{\varphi}_k+(F(x_k)-F(x^0))\varphi_k+B(x_k)d_k(x_k)\\ &-B(x^0)d_k(x^0)+(C(x_k)-C(x^0))\varepsilon_{k+1},\end{aligned}$$

则有 $\tilde{\varphi}_k\xrightarrow[k\to\infty]{}0$，所以 $\|u_k-u_k^0\|\xrightarrow[k\to\infty]{}0$，并且成立式(6.3.7)．

§6.4 连续时间系统的参数估计

我们在第五章中给出了连续时间随机逼近算法的收敛性定理. 这些结果可用来研究连续时间系统的参数估计问题，得到和§6.2、6.3相类似的结果.

设 $x^0\in R^l$ 是未知参数，$f(x^0)=0$. 设 x_t 是 t 时刻对 x^0 的估计，它的方程在以后给出.

设实际系统为

$$d\varphi_t=A(x_t)\varphi_t\,dt+B(x_t)\eta_t\,dt, \tag{6.4.1}$$

状态 φ_t 是完全观测的. 对参数 x^0 的信息通过量测方程

$$dy_t=Q(t,\ x_t,\ \varphi_t)dt+H(x_t)dw_t+\nu_t\,dt \tag{6.4.2}$$

来获得. $A(\cdot)$, $B(\cdot)$, η_t, $Q(t,\ x_t,\ \varphi_t)$及量测噪声

$$H(x_t)dw_t+\nu_t\,dt$$

都不事先给出.

我们要用到下面条件.

B1 设 M 闭, $M\subset R^l$, $x^0\in \overset{\circ}{M}$, 对任一 $x\in M$, $A(x)$稳定.

B2 $\{\eta_t,\ \mathscr{F}_t\}$ 有界, $\|\eta_t\|\leqslant\eta$, 且 $E\eta^2<\infty$, $\mathscr{F}_t\uparrow$.

B3 $A(x)$, $B(x)$, $H(x)$在 M 内满足 Lipschitz 条件.

B4 对 $x\in M$ 及 φ, $Q(t,\ x,\ \varphi)$满足局部 Lipschitz 条件，并且 Lipschitz 系数不依赖于 t.

B5 存在极限

$$\lim_{t\to\infty}EQ(t,\ x,\ \varphi_t(x))\triangleq f(x),\ \forall x\in M,$$

并且在 M 上，$f(x)$满足 Lipschitz 条件. 对固定的 $x\in R^l$, $\varphi_t(x)$是下面方程的解：

$$d\varphi_t(x)=A(x)\varphi_t(x)dt+B(x)\eta_t\,dt,\ \varphi_0(x)=0. \tag{6.4.3}$$

B6 存在二次连续可微函数 $v(\cdot)$: $R^l\to R$, 使

$$\sup_{\substack{0<|x-x^0|\\ x\in M}} f^\tau(x)v_x(x)<0, \tag{6.4.4}$$

并且存在 $x^*\in R^l$ 及 $c_0>0$ 使

$$v(x^*)<\inf_{\|x\|=c_0} v(x), \tag{6.4.5}$$

$$\{x: \|x\|\leqslant c_0\}\subset \mathring{M},\ x^*\in\{x: \|x\|<c_0\}. \tag{6.4.6}$$

B7 $(w_t,\ \mathscr{F}_t)$是标准 Wiener 过程. 除了一个可能的例外零测集外，对每个ω，$(\nu_t,\ \mathscr{F}_t)$满足以下条件：当 $k\to\infty$ 时，$t_k\to\infty$，x_{t_k} 收敛时，则

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\int_{t_k}^{t} a_s\nu_s\,\mathrm{d}s=0,\ \forall t\in[t_k,\ s(t_k,\ T)] \tag{6.4.7}$$

B8 $a_s>0,\ a_s\xrightarrow[s\to\infty]{}0,\ \int_0^\infty a_s\,\mathrm{d}s=\infty,\ \int_0^\infty a_s^2\,\mathrm{d}s<\infty.$

对 x^0 的估计 x_t 定义如下：

$$\begin{aligned}x_t^{i+1}&=x^*+\int_{\sigma_i}^{t} a_s\,\mathrm{d}y_s\\&=x^*+\int_0^t a_s[Q(s,\ x_s^{i+1},\ \varphi_s)+\nu_s]I_{[\sigma_i<s]}\mathrm{d}s\\&\quad+\int_0^t a_sH(x_s^{i+1})I_{[\sigma_i<s]}\,\mathrm{d}w_s,\ t<\sigma_{i+1},\end{aligned} \tag{6.4.8}$$

$$\sigma_0=0,\ \sigma_{i+1}=\begin{cases}\inf\{t: t>\sigma_i,\ \ x_t^{i+1}\overline{\in}\mathring{M}\},\\ \infty,\ \text{如果 } x_t^{i+1}\in\mathring{M},\ \forall t>\sigma_i,\ i=0,\ 1,\ \cdots,\end{cases} \tag{6.4.9}$$

$$x_t=\sum_{i=1}^{\infty}x_t^iI_{[\sigma_{i-1}\leqslant t<\sigma_i]}. \tag{6.4.10}$$

用定理1.3.3知方程(6.4.1)及(6.4.8)在区间$[\sigma_i,\ \sigma_{i+1})$上有唯一解$(x_t^{i+1},\ \varphi_t)$，爆炸时间比 σ_{i+1} 大. 由于 $A(x_t)$ 及 $B(x_t)\eta$ 有界，所以式(6.4.1)对 φ_t 线性，因而没有爆炸. 我们现在来证 $\sigma_i\to\infty$. 反设 $\sigma_i\xrightarrow[i\to\infty]{}\sigma<\infty$，那么 $\sigma_{i+1}-\sigma_i\xrightarrow[i\to\infty]{}0$. 这就导致下面的矛盾：

$$\begin{aligned}0<\mathrm{d}(x^*,\ (\mathring{M})^C)&\leqslant\|x_{\sigma_{i+1}}-x^*\|\\&\leqslant\left\|\int_{\sigma_i}^{\sigma_{i+1}}a_s[Q(s,\ x_s^{i+1},\ \varphi_s)+\nu_s]\,\mathrm{d}s\right\|\end{aligned}$$

$$+\left\|\int_{\sigma_i}^{\sigma_{i+1}} a_s H(x_s^{i+1})\mathrm{d}w_s\right\| \xrightarrow[i\to\infty]{} 0.$$

所以式(6.4.10)给出的 x_t 在 $(0, \infty)$ 上有定义.

定理 6.4.1 设 B1～B8 成立，下面所涉及的条件概率正则。又设每当 x_{t_k} 收敛到 M 的内点时，$t_k \xrightarrow[k\to\infty]{} \infty$，相应的 $\|\varphi_{t_k}\|$ 必有界(除了可能的测度为零的例外集)，那么由式(6.4.10)定义的 $x_t \xrightarrow[t\to\infty]{} x^0$.

证明 记

$$\varepsilon_t = Q(t, x_t, \varphi_t) - f(x_t), \tag{6.4.11}$$

设 ω 固定，$x_{t_k} \xrightarrow[k\to\infty]{} x \in \overset{\circ}{M}$. 记

$$\varepsilon_{t,1} = Q(t, x_t, \varphi_t) - Q(t, x, \varphi_t(x)), \tag{6.4.12}$$

$$\varepsilon_{t,2} = Q(t, r, \varphi_t(r)) - \mathrm{E}Q(t, r, \varphi_t(r)), \tag{6.4.13}$$

$$\varepsilon_{t,3} = \mathrm{E}Q(t, x, \varphi_t(x)) - f(x_t) \tag{6.4.14}$$

$$\varepsilon_t^2 = Q(t, x, \varphi_t(x)) - \mathrm{E}Q(t, x, \varphi_t(x)) - \varepsilon_{t,2}, \tag{6.4.15}$$

$r \in \overset{\circ}{M}$, r 的分量为有理数.

从 $A(x)$ 的稳定性及连续性知

$$\|e^{A(x)t}\| \leqslant c_0 e^{-\lambda t},\ \lambda > 0,\ \forall x \in M, \tag{6.4.16}$$

由此知

$$\|\varphi_t(x)\| \leqslant C_0 \int_0^t e^{-\lambda(t-s)} b\eta\, \mathrm{d}s \leqslant \frac{c_0 b\eta}{\lambda},\ b = \max_{x\in M}\|B(x)\|, \tag{6.4.17}$$

注意到

$$\begin{aligned}\mathrm{d}(\varphi_t(x) - \varphi_t(y)) = {} & A(x)(\varphi_t(x) - \varphi_t(y))\mathrm{d}t \\ & + (A(x) - A(y))\varphi_t(y)\mathrm{d}t \\ & + (B(x) - B(y))\eta_t\,\mathrm{d}t,\end{aligned}$$

$$\begin{aligned}\varphi_t(x) - \varphi_t(y) = \int_0^t e^{A(x)(t-s)}[& (A(x) - A(y))\varphi_s(y) \\ & + (B(x) - B(y))\eta_s]\mathrm{d}s.\end{aligned}$$

由式(6.4.16)、(6.4.17)知，$\varphi_t(x)$ 对 $t \in [0, \infty)$ 一致地对 x 连续。由条件 B4、(6.4.17)及 $\mathrm{E}\eta < \infty$ 知，$Q(t, x, \varphi_t(x))$ 和 $\mathrm{E}Q(t, x,$

$\varphi_t(x)$)都对 $t\in[0,\infty)$一致地对 x 连续. 所以对任意 $\varepsilon>0$, 必可找 r, 使 $\|\varepsilon_t^0\|<\varepsilon$, $\forall t\in[0,\infty)$. 这样, 我们只要验证

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t} a_s\varepsilon_{s,j}\,\mathrm{d}s\right\|=0,\ \forall t\in[t_k,\ s(t_k,\ T)],$$
$$j=1,\ 2,\ 3 \qquad (6.4.18)$$

定理的结论就得自定理 5.2.1.

第1步 我们来证

$$\int_0^\infty a_s e_{s,2}\,\mathrm{d}s<\infty \quad \text{a.s.}, \qquad (6.4.19)$$

并且例外集可选成与 r 无关.

由于 r 可列, 所以只要对固定的 r 证明式(6.4.19).

用 η_{st} 表示 $\{\eta_\lambda,\ s\leqslant\lambda\leqslant t\}$ 产生的 σ-代数. 用 $P_t^{\eta_{0s}}(\mathrm{d}y)$表示给定 η_{0s} 后, η_λ, $s\leqslant\lambda\leqslant t$ 的条件分布, 而用 $P_s(\mathrm{d}y)$表示 η_λ, $0\leqslant\lambda\leqslant s$ 的分布. 下面约定, 用 η_s、$\varphi_s(r)$表示基本概率空间中的随机量, 而用 η_s^y, $\varphi_s^y(r)$表示它们的样本, 那么

$$\begin{aligned}
&\|\mathrm{E}(\varepsilon_{t+s,2}|\eta_{0s})\|\\
&=\left\|\int P_t^{\eta_{0s}}(\mathrm{d}y)Q\left(t+s,\ r,\ e^{A(r)t}\varphi_s(r)+\int_s^t e^{A(r)(t-\mu)}B(r)\eta_\mu^y\,\mathrm{d}\mu\right)\right.\\
&\quad-\int P_s(\mathrm{d}y)P_t^{\eta_{0s}}(\mathrm{d}y)Q\left(t+s,\ r,\ e^{A(r)t}\varphi_s^y(r)\right.\\
&\quad\left.\left.+\int_s^t e^{A(r)(t-\mu)}B(r)\eta_\mu^y\,\mathrm{d}\mu\right)\right\|\\
&=\left\|\int P_s(\mathrm{d}y)P_t^{\eta_{0s}}(\mathrm{d}y)\Big[Q(t+s,\ r,\ e^{A(r)t}\varphi_s(r)\right.\\
&\quad+\int_s^t e^{A(r)(t-\mu)}B(r)\eta_\mu^y\mathrm{d}\mu)-Q(t+s,\ r,\ e^{A(r)t}\varphi_s^y(r)\\
&\quad\left.+\int_s^t e^{A(r)(t-\mu)}B(r)\eta_\mu^y\,\mathrm{d}\mu)\Big]\right\|.
\end{aligned}$$

由条件 B4、式(6.4.16)、(6.4.17)及上式知

$$\begin{aligned}
&\|\mathrm{E}(\varepsilon_{t+s,2}|\eta_{0s})\|\\
&\leqslant\int P_s(\mathrm{d}y)K\|e^{A(r)t}\varphi_s(r)-e^{A(r)t}\varphi_s^y(r)\|\\
&\leqslant C_0Ke^{-\lambda t}(\|\varphi_s(r)\|+\mathrm{E}\|\varphi_s(r)\|)
\end{aligned}$$

$$\leqslant c_1 e^{-\lambda t}, \quad \mathrm{E} c_1^2 < \infty,$$

$$\mathrm{E}\|\varepsilon_{t,2}\| = \int P_t(\mathrm{d}y)\|Q(t, r, \varphi_t^y(r)) - \int P_t(\mathrm{d}z) Q(t, r, \varphi_t^z(r))\|$$

$$= \int P_t(\mathrm{d}y)\|\int P_t(\mathrm{d}z)[Q(t, r, \varphi_t^y(r)) - Q(t, r, \varphi_t^z(r))]\|$$

$$\leqslant 2K\mathrm{E}\|\varphi_t(r)\|,$$

这里 K 是 Lipschitz 常数，c_i 可依赖于 ω.

这样从 $\mathrm{E}\eta^2 < \infty$ 及式(6.4.17)知

$$\|\mathrm{E}(\varepsilon_{s,2}\varepsilon_{s+t,2}^{\tau})\| = \|\mathrm{E}[\varepsilon_{s,2}\mathrm{E}(\varepsilon_{s+t,2}^{\tau}|\eta_{0s})]\| \leqslant c_3 e^{-\lambda t},$$

所以

$$\mathrm{E}\left\|\int_0^\infty a_s \varepsilon_{t,2}\,\mathrm{d}s\right\|^2 = 2\mathrm{E}\int_0^\infty\int_0^\infty a_s a_{s+t}\varepsilon_{s,2}^{\tau}\varepsilon_{s+t,2}\,\mathrm{d}s\,\mathrm{d}t$$

$$\leqslant c_3\int_0^\infty\int_0^\infty a_s a_{s+t} e^{-\lambda t}\,\mathrm{d}t\,\mathrm{d}s$$

$$\leqslant \frac{c_3}{2}\left[\int_0^\infty\int_0^\infty a_s^2 e^{-\lambda t}\,\mathrm{d}s\,\mathrm{d}t + \int_0^\infty\int_0^\infty a_{s+t}^2 e^{-\lambda t}\,\mathrm{d}s\,\mathrm{d}t\right] < \infty,$$

由此得式(6.4.19)，所以式(6.4.18)对 $j=2$ 成立.

第 2 步 我们来证明对任意充分大的 k，可取 T_k 足够小使

$$\|\varphi_s\| \leqslant c_4 + c_5\Delta_s,$$

$$\Delta_s \triangleq \max_{t_k < t < s}\|x_t - x\|, \ \forall s \in [t_k, s(t_k, T_k)], \qquad (6.4.20)$$

注意到 x_s, φ_{t_k} 及 η_s 分别对 s, k 及 s 有界，从

$$\varphi_t = e^{A(x)(t-t_k)}\varphi_{t_k} + \int_{t_k}^t e^{A(x)(t-s)}[(A(x_s) - A(x))\varphi_s + B(x_s)\eta_s]\,\mathrm{d}s,$$

由式(6.4.16)及条件 B3 知

$$\|\varphi_t\| \leqslant c_6 e^{-\lambda(t-t_k)} + c_0 b\eta\int_{t_k}^t e^{-\lambda(t-s)}\mathrm{d}s + c_7\int_{t_k}^t e^{-\lambda(t-s)}\Delta_s\|\varphi_s\|\,\mathrm{d}s$$

$$\leqslant c_8 + c_9\int_{t_k}^t e^{-\lambda(t-s)}\Delta_s\|\varphi_s\|\mathrm{d}s.$$

记 $h_t = e^{\lambda t}\|\varphi_t\|$，由定理 1.3.4，从上式得

$$h_t \leqslant c_8 e^{\lambda t} + c_8 c_9 \int_{t_k}^{t} \Delta_s \exp\left[\lambda s + c_9 \int_{s}^{t} \Delta_\mu \, \mathrm{d}\mu\right] \mathrm{d}s,$$

因此

$$\|\varphi_t\| \leqslant c_8 + c_8 c_9 \Delta_t \int_{t_k}^{t} e^{-(\lambda - C_9 \Delta_t)(t-s)} \mathrm{d}s$$

$$= c_8 + \frac{c_8 c_9 \Delta_t}{\lambda - c_9 \Delta_t}(1 - e^{-(\lambda - C_9 \Delta_t)t}). \qquad (6.4.21)$$

因为 $x_{t_k} \xrightarrow[k\to\infty]{} x \in \mathring{M}$，所以对充分大的 k 及充分小的 T_k，对 $t \in [t_k, s(t_k, T_k)]$ 算法没有截尾，从而 x_t 在 $[t_k, s(t_k, T_k)]$ 上连续。于是

$$\lim_{T_k \to 0} \lim_{k \to \infty} \Delta_t = 0, \ \forall t \in [t_k, s(t_k, T_k)].$$

所以对充分大的 k，总可取小的 T_k 使

$$\lambda - c_9 \Delta_{s(t_k, T_k)} \geqslant \frac{\lambda}{2}, \qquad (6.4.22)$$

那么从式(6.4.21)、(6.4.22)立即得出式(6.4.20)。

第3步 我们现在来证存在不依赖于 k 的 $c_{10}>0$，$c_{11}>0$ 及 T_0，对一切 $T \leqslant T_0$ 及充分大的 k 成立

$$\Delta_t \leqslant c_{10} T, \ \|\varphi_t\| \leqslant c_{11}, \ \forall t \in [t_k, s(t_k, T)]. \qquad (6.4.23)$$

注意到对充分大的 k

$$x_t = x_{t_k} + \int_{t_k}^{t} a_s [Q(s, x_s, \varphi_s) + \nu_s] \mathrm{d}s + \int_{t_k}^{t} a_s H(x_s) \mathrm{d}\omega_s, \ t \in [t_k, s(t_k, T_k)]. \qquad (6.4.24)$$

从 Lipschitz 条件及 $\{x_t\}$ 的有界性知

$$\|Q(s, x_s, \varphi_s)\| \leqslant c_{12} + c_{13} \|\varphi_s\|. \qquad (6.4.25)$$

这和式(6.4.24)一起导致

$$\Delta_t \leqslant \|x_{t_k} - x\| + c_{12} T + c_{13} \int_{t_k}^{t} a_s \|\varphi_s\| \mathrm{d}s + \int_{t_k}^{t} a_s \nu_s \mathrm{d}s + \int_{t_k}^{t} a_s H(x_s) \mathrm{d}\omega_s, \ \forall t \in [t_k, s(t_k, T)], \ \forall T \leqslant T_k. \qquad (6.4.26)$$

由式(6.4.20)及式(6.4.7), 并注意到

$$\int_0^\infty a_s H(x_s) \mathrm{d}\omega_s < \infty \quad \text{a.s.},$$

从式(6.4.26)知

$$\Delta_t \leqslant c_{12}T + c_{13}T(c_4 + c_5\Delta_t) + g_{kt}, \tag{6.4.27}$$

这里 $$\lim_{T\to 0} \limsup_{k\to\infty} \|g_{k,t}\| = 0.$$

所以,

$$\Delta_t \leqslant c_{14}T + c_{15}T\Delta_t,$$

当 $T_k < 1/c_{15}$ 时,

$$\Delta_t \leqslant \frac{c_{14}T}{1-c_{15}T}, \ \forall t \in [t_k, s(t_k, T)], \ \forall T \leqslant T_k. \tag{6.4.28}$$

为使式(6.4.28)成立, 要求 T_k 小, 使 i) $[t_k, s(t_k, T_k)]$ 上无截尾, ii) 式(6.4.22)成立, iii) $T_k < 1/c_{15}$. 记

$$T(k) = \max\{T: \lambda - c_9\Delta_{s(t_k, T)} \geqslant \lambda/2, \text{在}[t_k, s(t_k, T)]\text{上无截尾}\} \tag{6.4.29}$$

在上面的分析中, 取

$$T_k = \left(T(k) \wedge \frac{1}{c_{15}}\right), \tag{6.4.30}$$

并取 $T_0 > 0$,

$$T_0 < \left(\frac{1}{2c_{15}} \wedge \frac{\lambda}{8c_9c_{14}} \wedge \frac{d}{2c_{14} + dc_{15}}\right), \tag{6.4.31}$$

这里 $$d = d(x, \partial M).$$

我们来证, 对一切充分大的 k, $T(k) \geqslant T_0$. 显然, 我们只要考察 $T(k) < \infty$ 情形.

由于 x_t 连续, 我们只要考察 i) $\lambda - c_9\Delta_{s(t_k, T(k))} = \dfrac{\lambda}{2}$ 或 ii) 在 $s(t_k, T(k))$ 时, 发生截尾.

当 $T(k) \geqslant \dfrac{1}{c_{15}}$ 时, 由式(6.4.31)知 $T(k) > T_0$. 所以我们只要考察 $T(k) < \dfrac{1}{c_{15}}$ 的情形.

当 ii)成立且 $T(k)<T_0$ 时，那么在 $[t_k,\ s(t_k,\ T(k)))$ 上没有截尾，并且 $\lambda-c_9\Delta_{s(t_k,T(k))}\geqslant\frac{\lambda}{2}$，$T(k)<\frac{1}{c_{15}}$. 因此从式(6.4.28)知

$$\Delta_t\leqslant\frac{c_{14}T(k)}{1-c_{15}T(k)},\ \forall t\in[t_k,\ s(t_k,\ T(k))),$$

由式(6.4.31)还知

$$\Delta_t<\frac{c_{14}T_0}{1-c_{15}T_0}\leqslant\frac{d}{2}.$$

这表明在 $s(t_k,\ T(k))$ 处没有截尾，因此和 ii) 成立的假设相矛盾. 所以只能 $T(k)\geqslant T_0$ 或 ii)不成立. $T(k)\geqslant T_0$ 正是我们想要证的，而当 ii)不成立时，我们只要讨论 i)成立的情形就够了.

设 i)成立且 $T(k)<T_0$. 那么从式(6.4.28)，及式(6.4.31)知

$$\Delta_{s(t_k,T(k))}\leqslant\frac{c_{14}T(k)}{1-c_{15}T(k)}<\frac{c_{14}T_0}{1-c_{15}T_0}<2C_{14}T_0.$$

这和 i)及式(6.4.31)一起得

$$\begin{aligned}\frac{\lambda}{2}&=\lambda-c_9\Delta_{s(t_k,T(k))}>\lambda-2c_9c_{14}T_0\\&>\lambda-\frac{\lambda}{4}=\frac{3\lambda}{4}.\end{aligned}$$

所得矛盾表明对一切充分大的 k，$T(k)\geqslant T_0$. 由式(6.4.30)便知 $T_k\geqslant T_0$. 从而由式(6.4.28)、(6.4.20)就得式(6.4.23).

第4步 我们现对 $j=1$ 及 3 来证式(6.4.18).

记
$$\tilde{\varphi}_t=\varphi_t-\varphi_t(x).$$

因为 $\|\varphi_{t_k}\|<c$，由式(6.4.17)知 $\|\tilde{\varphi}_{t_k}\|\leqslant c+\frac{c_0b\eta}{\lambda}$.

由条件 B2、B3、和式(6.4.16)、(6.4.23)，从方程

$$\begin{aligned}d\tilde{\varphi}_t=&A(x)\tilde{\varphi}_t\,dt+(A(x_t)-A(x))\varphi_t\,dt\\&+(B(x_t)-B(x))\eta_t\,dt\end{aligned}$$

知对一切充分大的 K 有

$$\|\tilde{\varphi}_t\|\leqslant c_{16}e^{-\lambda(t-t_k)}+c_{17}T,\ \forall t\in[t_k,\ s(t_k,\ T)],\ \forall T\leqslant T_0.\tag{6.4.32}$$

由条件 B3、式(6.4.23)及式(6.4.32)就有

$$\lim_{T\to 0}\limsup_{k\to\infty}\left\|\frac{1}{T}\int_{t_k}^{t} a_{s\varepsilon_{s,1}}\mathrm{d}s\right\|$$
$$\leqslant K\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\int_{t_k}^{t} r_s(\Delta_s+\|\tilde{\varphi}_s\|)\mathrm{d}s$$
$$\leqslant K\lim_{T\to 0}\limsup_{k\to\infty}\left(c_{10}T+\frac{1}{T}\max_{t_k\leqslant s\leqslant t}a_s c_{10}\int_{t_k}^{t} e^{-\lambda(s-t_k)}\mathrm{d}s+c_{17}T\right)$$
$$=0,$$
$$\forall t\in[t_k,\ s(t_k,\ T)].$$

这就证明了式(6.4.18)对 $j=1$ 成立.

最后由条件 B5、$f(\cdot)$的连续性及式(6.4.23)知

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t} a_s e_{s,3}\mathrm{d}s\right\|$$
$$\leqslant\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t} a_s(f(x)-f(x_s))\mathrm{d}s\right\|$$
$$+\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t}(\mathrm{E}Q(s,\ x,\ \varphi_s(x))-f(x))r_s\mathrm{d}s\right\|$$
$$=0,\quad \forall t\in[t_k,\ s(t_k,\ T)].$$

这就对 $j=3$ 证明了式(6.4.18).　□

注 4.4.1　在条件 B1 中，M 可放宽为无界集，这时算法(6.4.8)、(6.4.9)就应从固定界截尾改为变界截尾，如(5.2.8)～(5.2.11)定义的那样.

注 4.4.2　和§6.3类似地可以讨论连续时间的适应控制问题. 把定理 6.4.1 用来求二次指标下的适应控制.

§6.5　非线性系统的参数估计

在上一节我们所讨论的实际系统(6.4.1)对 φ_t 是线性的. 现在我们来讨论非线性系统，噪声也较系统(6.4.1)中的更一般.

设实际系统为

$$\mathrm{d}\varphi_t=A(x_t)\varphi_t\mathrm{d}t+g(x_t,\ \varphi_t)\mathrm{d}t+B(x_t)u_t\mathrm{d}t+D(x_t)\zeta_t\mathrm{d}t+C(\varphi_t)\mathrm{d}w_t,\tag{6.5.1}$$

$\varphi_t\in R^m$，初值φ_0是确定性矢量．量测方程仍和式(6.4.2)一样

$$dy_t=Q(t,\ x_t,\ \varphi_t)dt+H(x_t)d\omega_t+\nu_t dt,\ x_t\in R^l. \quad (6.5.2)$$

和上一节一样，我们要估计$x^0\in R^l$，$f(x^0)=0$，x_t是t时刻对x^0的估计，$Q(t,\ x,\ \varphi)$和x^0的联系将在下面的条件中给出．

我们将用到下面条件．

C1　$M\subset R^l$，M闭，$x^0\in\mathring{M}$．用$z=\{z_t\}$表示任意取值于M的连续函数．用$\Phi_{t,s}(z^-)$表转移函数阵

$$\frac{d\Phi_{t,s}(z)}{dt}=A(z_t)\Phi_{t,s}(z),\ \Phi_{s,s}(z_0)=I. \quad (6.5.3)$$

设存在不依赖于z的常数$c>0$，$\lambda>0$使

$$\|\Phi_{t,s}(z)\|\leqslant ce^{-\lambda(t-s)},\quad \forall t\geqslant s\geqslant 0. \quad (6.5.4)$$

C2　用$\nu_{0,t}$表示$\sigma\{\nu_s,\ 0\leqslant s\leqslant t\}$，$\mathscr{F}_t\triangleq\sigma\{\nu_{0t},\zeta_{0t},\ w_{0t}\}$，$(w_t,\mathscr{F}_t)$是标准Wiener过程$\{\zeta_t\}$、$\{\omega_t\}$和$\{\nu_t\}$相互独立，$\|\nu_t\|\leqslant\nu$，$E\nu^2<\infty$，$\|\zeta_t\|<\zeta$，$\zeta$为常数．除了可能存在的零测集，对一切$\omega$，只要$x_{t_k}$当$t_k\to\infty$时收敛，就有

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t}r_s\nu_s ds\right\|=0,\ \forall t\in[t_k,\ s(t_k,\ T)], \quad (6.5.5)$$

这里

$$r_s>0,\ \int_0^\infty r_s ds=\infty,\ \int_0^\infty r_s^2 ds<\infty,\ r_s\xrightarrow[s\to\infty]{}0, \quad (6.5.6)$$

$$\int_t^{s(t,T)} r_s ds=T. \quad (6.5.7)$$

C3　$u_t\in\mathscr{F}_t$，$\|u_t\|\leqslant u$，u为常数．

C4　存在有界集$M_1\subset R^m$，使

$$C(\varphi)=0,\quad \forall\varphi\overline{\in}M_1.$$

$\|g(x,\ \varphi)\|\leqslant g$，$\forall x\in M$，$\forall\varphi\in R^m$，并且$g(x,\ \varphi)$及$C(\varphi)$满足Lipschitz条件

$$\|g(x,\ \varphi)-g(y,\ \psi)\|\leqslant k_1\|x-y\|+k_2\|\varphi-\psi\|, \quad (6.5.8)$$

$$\|C(\varphi)-C(\psi)\|\leqslant k_3\|\varphi-\psi\|,$$

设 k_2、k_3 和式(6.5.4)中的 λ, c 满足以下关系

$$\lambda-2c^2\left(\frac{k_2^2}{\lambda}+k_3^2\right)>0, \tag{6.5.9}$$

(由此知存在 $\delta>0$ 使

$$\lambda-(2+\delta)c^2\left(\frac{k_2^2}{\lambda}+k_3^2\right)\triangleq\rho>0\Big).$$

C5　$A(x)$, $B(x)$, $D(x)$, $H(x)$, $Q(t, x, \varphi)$对 φ 及 $x\in M$ Lipschitz 连续. 对 $Q(t, x, \varphi)$的 Lipschitz 常数不依赖于 t.

C6　设

$$\begin{cases} d\varphi_t(x)=A(x)\varphi_t(x)dt+g(x, \varphi_t(x))dt+B(x)u_t dt \\ \qquad\qquad +D(x)\zeta_t dt+C(\varphi_t(x))dw_t, \\ \varphi_0(x)=0, \end{cases} \tag{6.5.10}$$

对任意 $x\in M$, 存在 $f(x)$使

$$\lim_{t\to\infty}EQ(t, x, \varphi_t(x))\triangleq f(x), \tag{6.5.11}$$

并且在 M 上, $f(\cdot)$为 Lipschitz 连续.

C7　存在二次连续可微函数 $v(\cdot)$: $R^l\to R$, 使

$$\sup_{\substack{0<\|x-x^0\|,\\ x\in M}} f^\tau(x)v_x(x)<0, \tag{6.5.12}$$

还存在 $x^*\in R^m$ 及 $c_0>0$ 使

$$v(x^*)<\inf_{\|x\|=C_0}v(x), \tag{6.5.13}$$

$$\{x: \|x\|\leqslant C_0\}\subset\mathring{M},\ x^*\in\{x: \|x\|<c_0\}. \tag{6.5.14}$$

对 x^0 的估计 x_t 由下面的算法给出

$$x_t^{i+1}=x^*+\int_{\sigma_i}^t r_s dy_s=x^*+\int_0^t r_s[Q(s, x_s^{i+1}, \varphi_s)+\nu_s]I_{[\sigma_i<s]}ds$$
$$+\int_0^t r_s H(x_s^{i+1})I_{[\sigma_i<s]}dw_s, \tag{6.5.15}$$

$$\sigma_0=0,\ \sigma_{i+1}=\begin{cases}\inf\{t: t>\sigma_i,\ x_t^{i+1}\overline{\in}\mathring{M}\},\\ \infty,\ 如果\ x_t^{i+1}\in\mathring{M},\ \forall t>\sigma_i,\ i=0, 1, \cdots\end{cases}$$

$$x_t=\sum_{i=1}^{\infty}x_t^i I_{[\sigma_{i-1}\leqslant t<\sigma_i]}. \qquad (6.5.16)$$

用 Lipschitz 条件知，对式(6.5.1)及式(6.5.16)存在唯一强解.

引理 6.5.1 除了式(6.5.5)～(6.5.7)及式(6.5.9)外，设成立条件 C1～C5．那么存在常数 μ, 使

$$\|\varphi_t\|\leqslant\mu,\ \forall t\geqslant 0.$$

证明 设 $\psi_{t,0}=\Phi_{t,0}(x)$, 即

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_{t,0}=A(x_t)\psi_{t,0},\quad \psi_{0,0}=I. \qquad (6.5.17)$$

取 $c_1>(1\vee c\vee\|\varphi_0\|)$, $M_1\subset\{x:\ \|x\|\leqslant c_1\}$, 这里 c 在式(6.5.4)中给出，定义 $b=\max\limits_{x\in M}\|B(x)\|$,

$$\tau=\begin{cases}\inf\{t:\ \|\varphi_t\|\geqslant c_1^2+\dfrac{c}{\lambda}(g+bu+\max\limits_{x\in M}\|D(x)\|\zeta)\},\\ \infty,\ 若\ \|\varphi_t\|<c_1^2+\dfrac{c}{\lambda}(g+bu+\max\limits_{x\in M}\|D(x)\|\zeta),\ \forall t\geqslant 0.\end{cases}$$

对 $\tau<\infty$ 情形，必存在 $\tau_1<\tau$ 使 $\|\varphi_{\tau_1}\|=c_1$, 注意到 $\|\varphi_0\|<c_1$, 所以 $c_1<\|\varphi_t\|\leqslant c_1^2+\dfrac{c}{\lambda}(g+bu+\max\limits_{x\in M}\|D(x)\|\zeta)$, $\forall t\in(\tau_1,\tau]$. 这说明 $\varphi_t\bar{\in}M_1$, $\forall t\in(\tau_1,\tau]$.

这样我们就有

$$\begin{aligned}\varphi_\tau&=\psi_{\tau,\tau_1}\varphi_{\tau_1}+\int_{\tau_1}^{\tau}\psi_{\tau,s}(g(x_s,\ \varphi_s)\mathrm{d}s+B(x_s)u_s\mathrm{d}s+D(x_s)\zeta_s\mathrm{d}s)\\&\quad+\int_{\tau_1}^{\tau}\psi_{\tau,s}C(\varphi_s)\mathrm{d}w_s\\&=\psi_{\tau,\tau_1}\varphi_{\tau_1}+\int_{\tau_1}^{\tau}\psi_{\tau,s}[g(x_s,\ \varphi_s)+B(x_s)u_s+D(x_s)\zeta_s]\mathrm{d}s.\end{aligned}$$

由式(6.5.4)及(6.5.17)知

$$\begin{aligned}&c_1^2+\frac{c}{\lambda}(g+bu+\max_{x\in M}\|D(x)\|\zeta)\\&=\|\varphi_\tau\|\leqslant cc_1e^{-\lambda(\tau-\tau_1)}+c\int_{\tau_1}^{\tau}e^{-\lambda(\tau-s)}(g+bu+\max_{x\in M}\|D(x)\|\zeta)\mathrm{d}s\end{aligned}$$

$$<cc_1+\frac{c(g+bu+\max\limits_{x\in M}\|D(x)\|\zeta)}{\lambda}.$$

这就导致 $c_1<c$, 而这和 c_1 的定义相矛盾. 所以

$$\|\varphi_t\|<c_1^2+\frac{c}{\lambda}(g+bu+\max_{x\in M}\|D(x)\|\zeta)\triangleq\mu. \qquad \square$$

定理 6.5.1 在条件 C1～C7 下, 设下面所涉及的条件概率正则, 那么 $x_t\xrightarrow[t\to\infty]{}x^0$.

证明 固定 w, 设

$$x_{t_k}\xrightarrow[k\to\infty]{}x\in\mathring{M}, \quad d\triangleq d(x,\partial M)>0.$$

任给 $\varepsilon\in\left(0,\frac{d}{2}\right)$, 对充分大的 k, 有

$$\|x_{t_k}-x\|<\frac{\varepsilon}{2},$$

并且到截尾前有

$$x_t-x_{t_k}=\int_{t_k}^{t}r_s(Q(s,x_s,\varphi_s)+\nu_s)\mathrm{d}s+\int_{t_k}^{t}r_sH(x_s)\mathrm{d}w_s. \tag{6.5.18}$$

因为 $\int_0^\infty r_sH(x_s)\mathrm{d}w_s<\infty$, $Q(s,x_s,\varphi_s)$及 ν_s 有界, 存在 $\delta(\varepsilon)>0$, 当 $0<t-t_k<\delta(\varepsilon)$ 时,

$$\|x_t-x_{t_k}\|<\frac{\varepsilon}{2},$$

并且 $\delta(\varepsilon)$对一切充分大的 k 都一样.

取 n_ε 充分大, 使 $\frac{1}{n_\varepsilon}<\frac{\delta(\varepsilon)}{2}$. 那么对任一 k, 可找到正整数 m_k, 使 $0<\frac{m_k}{n_\varepsilon}-t_k<\delta(\varepsilon)$, 那么

$$\|x_{\frac{m_k}{n_\varepsilon}}-x_{t_k}\|<\frac{\varepsilon}{2}, \quad \|x_{\frac{m_k}{n_\varepsilon}}-x\|<\varepsilon.$$

总之, 任给 $\varepsilon>0$, 可找到正整数 n_ε 及 $\{m_k\}$, 使对一切充分大的 k 成立

$$\|x_{\frac{m_k}{n_\varepsilon}}-x\|<\varepsilon \tag{6.5.19}$$

设 $r\in R^l$, r 的分量都是有理数. 记

$$\varepsilon_{t,0}=Q(t,\ x_t,\ \varphi_t)-Q(t,\ r,\ \varphi_t),$$

$$\varepsilon_{t,1}=Q(t,\ r,\ \varphi_t)-\mathrm{E}(Q(t,\ r,\ \varphi_t)\,|\,\mathscr{F}_{\frac{m_k}{n_k}}),$$

$$\varepsilon_{t,2}=\mathrm{E}(Q(t,\ r,\ \varphi_t(r))\,|\,\mathscr{F}_{\frac{m_k}{n_k}})-\mathrm{E}Q(t,\ r,\ \varphi_t(r)),$$

$$\varepsilon_{t,3}=\mathrm{E}(Q(t,\ r,\ \varphi_t)\,|\,\mathscr{F}_{\frac{m_k}{n_k}})-\mathrm{E}(Q(t,\ r,\ \varphi_t(r))\,|\,\mathscr{F}_{\frac{m_k}{n_k}}),$$

$$\varepsilon_{t,4}=\mathrm{E}Q(t,\ r,\ \varphi_t(r))-f(x_t).$$

那么

$$Q(t,\ x_t,\ \varphi_t)-f(x_t)=\sum_{j=0}^{4}\varepsilon_{t,j}. \tag{6.5.20}$$

下面用类似于定理6.4.1的证明中所引进的符号, 用 $P_{st}^{\mathscr{F}_s}(\mathrm{d}y)$ 表示给定 $\mathscr{F}_s$ 后, 被积函数中出现的过程 x_λ, φ_λ, u_λ, ζ_λ, w_λ, $\Psi_{t+\lambda,\lambda}$, $s\leqslant\lambda\leqslant t$ 的联合分布, 并把 $\{x_\lambda,\ \varphi_\lambda,\ u_\lambda,\ \nu_\lambda,\ \zeta_\lambda,\ w_\lambda\}$ 简记为 $\{\eta_\lambda\}$. 把给定 $\mathscr{F}_s$ 后, $\{\eta_\lambda,\ 0\leqslant\lambda\leqslant t\}$ 的条件分布记为 $P_t^{\mathscr{F}_s}(\mathrm{d}y)$. 记

$$h(s,\ t)=\int_s^t\Psi_{t,\lambda}(g(x_\lambda,\ \varphi_\lambda)+B(x_\lambda)u_\lambda+D(x_\lambda)\zeta_\lambda)\mathrm{d}\lambda$$
$$+\int_s^t\Psi_{t,\lambda}C(\varphi_\lambda)\mathrm{d}w_\lambda,$$

那么

$$\varphi_t=\Psi_{t,s}\varphi_s+h(s,\ t).$$

因为

$$\mathrm{E}(Q(t+s,\ r,\ \varphi_{t+s})\,|\,\mathscr{F}_s)$$
$$=\int P_{s,t+s}^{\mathscr{F}_s}(\mathrm{d}y)Q(t+s,\ r,\ \Psi_{t+s,s}\varphi_s+h(s,\ t+s)) \tag{6.5.21}$$

的右端作为基本概率空间中的变量, 不再依赖于样本函数 η_λ, $\lambda\in[0,\ s]$, 所以

$$\int P_{s,t+s}^{\mathscr{F}_s}(\mathrm{d}y)Q(t+s,\ r,\ \Psi_{t+s,s}\varphi_s+h(s,\ t+s))$$

$$=\int P_{s,t+s}^{\mathscr{F}_{\frac{m}{n}}}(\mathrm{d}y)\int P_{s,t+s}^{\mathscr{F}_s}(\mathrm{d}y)Q(t+s,\ r,\ \Psi_{t+s,s}\varphi_s$$

$$+h(s,\ t+s)).\tag{6.5.22}$$

但式(6.5.21)的右端也可以在样本空间中考察，把它记为

$$\int P^{\mathscr{F}_s}_{s,t+s}(\mathrm{d}y)Q(t+s,\ r,\ \Psi_{t+s,s}\varphi^y_s+h(s,\ t+s)),$$

那么对 $\frac{m}{n}\leqslant s$ 有

$$\begin{aligned}&\mathrm{E}[(Q(t+s,\ r,\ \varphi_{t+s})\mid\mathscr{F}_{\frac{m}{n}})]\\&=\int P^{\mathscr{F}_{\frac{m}{n}}}_{s,t+s}(\mathrm{d}y)\int P^{\mathscr{F}_t}_{s,t+s}(\mathrm{d}y)Q(t+s,\ r,\ \Psi_{t+s,s}\varphi^y_s\\&\quad+h(s,\ t+s)).\end{aligned}\tag{6.5.23}$$

由式(6.5.21)～(6.5.23)知，对 $s\geqslant\frac{m}{n}$ 有如下估计：记

$$\xi_t=Q(t,\ r,\ \varphi_t)-\mathrm{E}(Q(t,\ r,\ \varphi_t)\mid\mathscr{F}_{\frac{m}{n}}),$$

由条件 C5 及式(6.5.4)有

$$\begin{aligned}&\|\mathrm{E}(\xi_{t+s}\mid\mathscr{F}_s)\|\\&\leqslant\left\|\int P^{\mathscr{F}_{\frac{m}{n}}}_{s,t+s}(\mathrm{d}y)\int P^{\mathscr{F}_t}_{s,t+s}(\mathrm{d}y)\right.\\&\quad\cdot[Q(t+s,\ r,\ \psi_{t+s,s}\varphi_s+h(s,\ t+s))\\&\quad\left.-Q(t+s,\ r,\ \Psi_{t+s,s}\varphi^y_s+h(s,\ t+s)]\right\|\\&\leqslant k\int P^{\mathscr{F}_{\frac{m}{n}}}_{s,t+s}(\mathrm{d}y)\int P^{\mathscr{F}_t}_{s,t+s}(\mathrm{d}y)\|\Psi_{t+s,s}\varphi_s-\Psi_{t+s,s}\varphi^y_s\|\\&\leqslant kce^{-\lambda t}(\|\varphi_s\|+\mathrm{E}(\|\varphi_s\|\mid\mathscr{F}_{\frac{m}{n}}))\\&\leqslant c_1e^{-\lambda t},\quad\forall t\geqslant s\geqslant 0,\end{aligned}\tag{6.5.24}$$

类似地；

$$\begin{aligned}&\mathrm{E}(\|\xi_t\|\mid\mathscr{F}_{\frac{m}{n}})\\&=\int P_t^{\mathscr{F}_{\frac{m}{n}}}(\mathrm{d}y)\left\|Q(t,\ r,\ \varphi^y_t)-\int P_t^{\mathscr{F}_{\frac{m}{n}}}(\mathrm{d}z)Q(t,\ r,\ \varphi^z_t)\right\|\\&=\int P_t^{\mathscr{F}_{\frac{m}{n}}}(\mathrm{d}y)\left\|\int P_t^{\mathscr{F}_{\frac{m}{n}}}(\mathrm{d}z)[Q(t,\ r,\ \varphi^y_t)-Q(t,\ r,\ \varphi^z_t)]\right\|\\&\leqslant 2k\mathrm{E}(\|\varphi_t\|\mid\mathscr{F}_{\frac{m}{n}})\leqslant c_2,\end{aligned}$$

c_1 和 c_2 都是常数。

由此及式(6.5.24)知

$$\mathrm{E}\left\|\int_{\frac{m}{n}}^{\infty} r_s\xi_s\,\mathrm{d}s\right\|^2 = 2\mathrm{E}\int_{\frac{m}{n}}^{\infty}\left(\int_0^{\infty} r_s r_{s+t}\xi_s^{\tau}\xi_{s+t}\mathrm{d}t\right)\mathrm{d}s$$

$$\leqslant 2\mathrm{E}\int_{\frac{m}{n}}^{\infty}\int_0^{\infty} r_s r_{s+t}\mathrm{E}(\|\xi_s^{\tau}\mathrm{E}(\xi_{t+s}|\mathscr{F}_s)\| |\mathscr{F}_{\frac{m}{n}})\mathrm{d}t\,\mathrm{d}s$$

$$\leqslant 2c_1c_2\int_{\frac{m}{n}}^{\infty}\int_0^{\infty} r_s r_{s+t}e^{-\lambda t}\,\mathrm{d}t/\mathrm{d}s$$

$$\leqslant c_1c_2\int_{\frac{m}{n}}^{\infty}\int_0^{\infty}(r_s^2+r_{s+t}^2)e^{-\lambda t}\mathrm{d}t\,\mathrm{d}s$$

$$\leqslant 2c_1c_2\int_0^{\infty} r_s^2\mathrm{d}s\int_0^{\infty} e^{-\lambda t}\,\mathrm{d}t<\infty.$$

所以 $\int_0^{\infty} r_s\xi_s\,\mathrm{d}s<\infty$ a.s..

因此，对任意 r 及正整数 r，

$$\int_{\frac{m}{n}}^{\infty} r_s\xi_s\,\mathrm{d}s \xrightarrow[m\to\infty]{} 0 \quad \text{a.s.}. \tag{6.5.25}$$

用 Ω' 表示 ω 集，在 Ω' 上式(6.5.5)成立，$\int_0^{\infty} r_s H(x_s)\,\mathrm{d}\omega_s<\infty$，式(6.5.25)对一切正整数 n 及具有有理分量的 r 都成立，那么 $P\Omega'=1$. 这样，可以认为在一开始固定的 $\omega\in\Omega'$. 由式(6.5.25)知

$$\int_{\frac{m_k}{n_k}}^{\infty} r_s e_{s,1}\,\mathrm{d}s \xrightarrow[k\to\infty]{} 0. \tag{6.5.26}$$

从式(6.5.10)可知

$$\varphi_{t+s}(r)=e^{A(r)t}\varphi_s(r)+q(s,\ t),$$

$$q(s,\ t)=\int_s^t e^{A(r)(t-\lambda)}[g(r,\ \varphi_\lambda(r))\mathrm{d}\lambda+B(r)u_\lambda\mathrm{d}\lambda$$

$$+D(r)\zeta_\lambda\mathrm{d}\lambda+C(\varphi_\lambda(r)\mathrm{d}w_\lambda].$$

与前面类似，不过现在用 $\{\eta_\lambda\}$ 代表 $\{\varphi_\lambda(r),\ u_\lambda,\ \zeta_\lambda,\ w_\lambda\}$，用 $P_{st}^{\eta_{0s}}(\mathrm{d}y)$ 表示给定 η_{0s} 后 $\varphi_\lambda(r)$，u_λ，ζ_λ，ω_λ，$s\leqslant\lambda\leqslant t$ 的联合分布，并用 $P_s(\mathrm{d}y)$ 表示 η_λ，$0\leqslant\lambda\leqslant s$ 的分布和式(6.5.24)类似地把 $\int P_{s,t+s}^{\eta_{0s}}(\mathrm{d}z)Q(t+s,\ r,\ e^{A(r)t}\varphi_s(r)+q(s,\ t+s))$ 当作基本概率空间

中的变量，而它在函数空间中的样本记为

$$\int P_{s,t+s}^{\eta_0,t}(\mathrm{d}z)Q(t+s,\ r,\ e^{A(r)t}\varphi_s^y(r)+q(s,\ t+s)).$$

为书写简便，记$s=\dfrac{m_k}{n_s}$，那么

$$\begin{aligned}e_{t+s,2}=&\mathrm{E}(Q(t+s,\ r,\ \varphi_{t+s}(r))\mid\eta_{0,s})-\mathrm{E}Q(t+s,\ r,\ \varphi_{t+s}(r))\\=&\int P_s(\mathrm{d}y)\int P_{s,t+s}^{\eta_0,t}(\mathrm{d}z)[Q(t+s,\ r,\ e^{A(r)t}\varphi_s(r)\\&+q(s,\ t+s))-Q(t+s,\ r,\ e^{A(r)t}\varphi_s^y(r)+q(s,\ t+s))].\end{aligned}$$

因此，由条件 C5 及式(6.5.4)就有

$$\begin{aligned}\|e_{t+s,2}\|\leqslant&kce^{-\lambda t}[E(\|\varphi_s(r)\|\mid\eta_{0,s})+\mathrm{E}\|\varphi_s(r)\|]\\\leqslant&c_3e^{-\lambda t}.\end{aligned}$$

所以，由上式及$r_t\xrightarrow[t\to\infty]{}0$，知

$$\begin{aligned}\left\|\int_{\frac{m_k}{n_s}}^{\infty}r_te_{t,2}\mathrm{d}t\right\|=&\left\|\int_0^{\infty}r_{t+\frac{m_k}{n_s}}e_{t+\frac{m_k}{n_s},2}\,\mathrm{d}t\right\|\\\leqslant&c_3\sup_{t\geqslant\frac{m_k}{n}}r_t\int_0^{\infty}e^{-\lambda t}\mathrm{d}t\xrightarrow[k\to\infty]{}0.\end{aligned}\qquad(6.5.27)$$

记
$$\tilde{\varphi}_t=\varphi_t(r)-\varphi_t,\qquad t\geqslant\frac{m_k}{n_s}.$$

注意到

$$\begin{aligned}\mathrm{d}\tilde{\varphi}_t=&A(r)\tilde{\varphi}_t\mathrm{d}t+(A(r)-A(x_t))\varphi_t\mathrm{d}t+(g(r,\ \varphi_t(r))\\&-g(x_t,\ \varphi_t)]\mathrm{d}t+B(r)-B(x_t))u_t\mathrm{d}t\\&+(D(r)-D(x_t))\zeta_t\mathrm{d}t+(c(\varphi_t(r))-c(\varphi_t))\mathrm{d}\omega_t,\end{aligned}$$

由条件 C4、C5 式(6.5.4)及引理 6.5.1 便知存在常数c_4使

$$\begin{aligned}\|\tilde{\varphi}_t\|\leqslant&ce^{-\lambda\left(t-\frac{m_k}{n_s}\right)}\|\tilde{\varphi}_{\frac{m_k}{n_s}}\|+c_4\int_{\frac{m_k}{n_s}}^{t}e^{-\lambda(t-s)}\|r-x_s\|\mathrm{d}s\\&+ck_2\int_{\frac{m_k}{n_s}}^{t}e^{-\lambda(t-s)}\|\tilde{\varphi}_s\|\mathrm{d}s+\left\|\int_{\frac{m_k}{n_s}}^{t}e^{A(r)(t-s)}\right.\\&\cdot[c(\varphi_s(r))-c(\varphi_s)]\mathrm{d}w_s,\end{aligned}\qquad(6.5.28)$$

这里k_2在式(6.5.8)中给出。

注意到初等不等式：

$$|xy|=\left|\frac{x}{\delta}\cdot\delta y\right|\leqslant\frac{2x^2}{\delta^2}+2\delta^2y^2,$$

所以对任意 $\delta>0$, 必可找到常数 $\alpha_1>0$, $\alpha_2>0$ 使

$$\left(\sum_{i=1}^{4}x_i\right)^2\leqslant\alpha_1x_1^2+\alpha_2x_2^2+(2+\delta)x_3^2+(2+\delta)x_4^2,$$

这里 x_i、x、y 为任意实数.

由条件 C4、引理 6.5.1 及 Schwarz 不等式从式(6.5.28)知,

$$\begin{aligned}\mathrm{E}(\|\tilde{\varphi}_t\|^2|\mathscr{F}_{\frac{m_k}{n_k}})\leqslant{}&c_5e^{-\lambda\left(t-\frac{m_k}{n_k}\right)}+c_6\int_{\frac{m_k}{n_k}}^{t}e^{-\lambda(t-s)}\\&\cdot\mathrm{E}(\|r-x_s\|^2|\mathscr{F}_{\frac{m_k}{n_k}})\mathrm{d}s+\frac{(2+\delta)c^2k_2^2}{\lambda}\\&\cdot\int_{\frac{m_k}{n_k}}^{t}e^{-\lambda(t-s)}\mathrm{E}(\|\tilde{\varphi}_s\|^2|\mathscr{F}_{\frac{m_k}{n_k}})\mathrm{d}s\\&+(2+\delta)k_3^2c^2\int_{\frac{m_k}{n_k}}^{t}e^{-2\lambda(t-s)}\mathrm{E}(\|\tilde{\varphi}_s\|^2|\mathscr{F}_{\frac{m_k}{n_k}})\mathrm{d}s,\end{aligned}\tag{6.5.29}$$

这里 c_5、c_6 为常数, k_2、k_3 为 C4 中的常数.

注意到 $Q(s,\ r,\ \varphi)$, ν_s 的有界性, $\int_0^\infty r_sH(x_s)\mathrm{d}w_s<\infty$, 由式(6.5.18)知, 如果 $\mathrm{d}(x_{\frac{m_k}{n_k}},\ \partial M)>\mathrm{d}$, 那么存在 $T>0$, 并且 T 不依赖于 k, 当 k 充分大时有

$$x_t-x_{\frac{m_k}{n_k}}=\int_{\frac{m_k}{n_k}}^{t}r_s(Q(s,\ x_s,\ \varphi_s)+\nu_s)\mathrm{d}s+\int_{\frac{m_k}{n_k}}^{t}r_sH(x_s)\mathrm{d}w_s,$$

$$\forall t\in\left[\frac{m_k}{n_k},\ s\left(\frac{m_k}{n_k},\ T\right)\right].\tag{6.5.30}$$

记

$$A_{m_k}=\left\{\omega:\ \mathrm{d}(x_{\frac{m_k}{n_k}},\ \partial M)>\frac{\mathrm{d}}{2}\right\}.\tag{6.5.31}$$

由式(6.5.30)知

$$\mathrm{E}(\|x_t-x_{\frac{m_k}{n_k}}\|^2I_{A_{m_k}}|\mathscr{F}_{\frac{m_k}{n_k}})\leqslant c_7T^2I_{A_{m_k}}+o(1)I_{A_{m_k}},$$

$$\forall t\in\left[\frac{m_k}{n_k},\ s\left(\frac{m_k}{n_k},\ T\right)\right],$$

或

$$\begin{aligned}&\mathrm{E}(\|x_t-r\|^2 I_{A_{m_k}}|\mathscr{F}_{\frac{m_k}{n_s}})\\&\leqslant 2\|x_{\frac{m_k}{n_\varepsilon}}-r\|^2 I_{A_{m_k}}+2c_7T^2I_{A_{m_k}}+o(1)I_{A_{m_k}},\qquad(6.5.32)\end{aligned}$$

这里 $o(1)$ 当 $k\to\infty$ 时趋于 0.

把式(6.5.32)代入式(6.5.29)，并用式(6.5.9)，记

$$h_t=e^{\lambda t}\mathrm{E}(\|\tilde{\varphi}_t\|^2 I_{A_{m_k}}|\mathscr{F}_{\frac{m_k}{n_s}}),\qquad(6.5.33)$$

那么就有

$$\begin{aligned}h_t\leqslant\ &c_5e^{\lambda\frac{m_k}{n}}I_{A_{m_k}}+c_7(\|x_{\frac{m_k}{n_s}}-r\|^2+T^2)e^{\lambda t}I_{A_{m_k}}\\&+(\lambda-\rho)\int_{\frac{m_k}{n_\varepsilon}}^t h_s\mathrm{d}s.\end{aligned}$$

取 $\varepsilon<T$，由式(6.5.19)，可取 r 充分接近 x，使 $\|x_{\frac{m_k}{n_s}}-r\|<T$，那么从上式知

$$h_t\leqslant c_5e^{\frac{\lambda m_k}{n_s}}I_{A_{m_k}}+c_8T^2e^{\lambda t}I_{A_{m_k}}+(\lambda-\rho)\int_{\frac{m_k}{n_s}}^t h_s\mathrm{d}s,$$

$$\forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right]\qquad(6.5.34)$$

从上式及定理 1.3.4 知

$$\begin{aligned}h_t\leqslant\ &c_5e^{\lambda\frac{m_k}{n_\varepsilon}}I_{A_{m_k}}+(\lambda-\rho)\int_{\frac{m_k}{n_s}}^t(c_5e^{\lambda\frac{m_k}{n_s}}+c_8T^2e^{\lambda s})\\&\times I_{A_{m_k}}e^{(\lambda-\rho)(t-s)}\mathrm{d}s,\end{aligned}$$

所以

$$\begin{aligned}&E(\|\tilde{\varphi}_t\|^2 I_{A_{m_k}}|\mathscr{F}_{\frac{m_k}{n_s}})\\&\leqslant c_5e^{-\lambda\left(t-\frac{m_k}{n_s}\right)}I_{A_{m_k}}+(\lambda-\rho)\int_{\frac{m_k}{n_s}}^t c_5e^{-\lambda\left(s-\frac{m_k}{n_\varepsilon}\right)-\rho(t-s)}\mathrm{d}s\,I_{A_{m_k}}\\&\quad+c_8(\lambda-\rho)T^2\int_{\frac{m_k}{n_s}}^t e^{-\rho(t-s)}\mathrm{d}sI_{A_{m_k}}\\&=c_5e^{-\lambda\left(t-\frac{m_k}{n_\varepsilon}\right)}I_{A_{m_k}}+c_6(e^{-\rho\left(t-\frac{m_k}{n_s}\right)}-e^{-\lambda\left(t-\frac{m_k}{n_s}\right)})I_{A_{m_k}}\end{aligned}$$

$$+\frac{c_8(\lambda-\rho)T^2}{\rho}\left(1-e^{-\rho\left(t-\frac{m_k}{n_s}\right)}\right)I_{A_{m_k}}$$

$$<c_5 e^{-\rho\left(t-\frac{m_k}{n_s}\right)}I_{A_{m_k}}+c_9T^2I_{A_{m_k}},\quad \forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right].$$

(6.5.35)

对我们所考察的固定的 ω, $x_{t_k}\to x$, 由式(6.5.19)及 $\varepsilon\in\left(0,\ \frac{\mathrm{d}}{2}\right)$, 从式(6.5.31)知对充分大的 k, 这个 $\omega\in A_{m_k}$. 因此对这个 ω, 由式(6.5.35)得

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\int_{\frac{m_k}{n_s}}^{t} r_s\mathrm{E}(\|\tilde{\varphi}_s\| \mid \mathscr{F}_{\frac{m_k}{n_s}})\mathrm{d}s$$

$$\leqslant\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\int_{\frac{m_k}{n_s}}^{t} r_s[\mathrm{E}(\|\tilde{\varphi}_s\|^2 \mid \mathscr{F}_{\frac{m_k}{n_s}})]^{\frac{1}{2}}\mathrm{d}s$$

$$\leqslant\lim_{T\to 0}\limsup_{k\to\infty}\left[\sup_{s>\frac{m_k}{n_s}} r_s\ \frac{1}{T}\int_{\frac{m_k}{n_s}}^{t}\sqrt{c_5}\ e^{-\frac{\rho}{2}\left(t-\frac{m_k}{n_s}\right)}\mathrm{p}s\right.$$

$$\left.+\sqrt{c_9}\int_{\frac{m_k}{n_s}}^{t} r_s\mathrm{d}s\right]=0,\quad \forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right].$$

由此及条件 C5 便知,

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{\frac{m_k}{n_s}}^{t} r_s e_{s,3}\,\mathrm{d}s\right\|=0,$$

$$\forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right].\qquad (6.5.36)$$

对任意 $\varepsilon>0$, 可选 r 离 x 充分近, 由条件 C5 知,

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{\frac{m_k}{n_s}}^{t} r_s\varepsilon_{s,0}\,\mathrm{d}s\right\|=0,$$

$$\forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right],\qquad (6.5.37)$$

由条件 C6 知 $\mathrm{E}Q(t,\ r,\ \varphi_t(r))\to f(r)$, 但 $f(x_{t_k})\to f(x)$, 所以

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{\frac{m_k}{n_s}}^{t} r_s e_{s,4}\,\mathrm{d}s\right\|<\varepsilon_1(r)\xrightarrow[r\to x]{} 0,$$

$$\forall t\in\left[\frac{m_k}{n_s},\ s\left(\frac{m_k}{n_s},\ T\right)\right].$$

这样，我们就证明了，对任意 $\varepsilon>0$，如果 $x_{t_k}\to x\in \mathring{M}$，则

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{\frac{m_k}{n_\varepsilon}}^{t} r_s(Q(s,\ x_s,\ \varphi_s)-f(x_s))\mathrm{d}s\right\|<\varepsilon,$$

$$\forall t\in\left[\frac{m_k}{n_\varepsilon},\ s\left(\frac{m_k}{n_\varepsilon},\ T\right)\right].$$

因为对 $x\in M$，$Q(s,\ x,\ \varphi)$ 及 $f(x)$ 有界，并且固定 $\varepsilon>0$ 后，m_k 是这样选择的(见式(6.5.18)以后)：

$$0<\frac{m_k}{n_\varepsilon}-t_k<\delta(\varepsilon)\xrightarrow[\varepsilon\to 0]{}0,$$

所以

$$\begin{aligned}&\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{\frac{m_k}{n_\varepsilon}} r_s(Q(s,\ x_s,\ \varphi_s)-f(x_s))\mathrm{d}s\right\|\\&\leqslant\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\sup_{s>t_k} r_s\cdot\delta(\varepsilon)\\&\quad\cdot\sup_{\substack{x\in M\\|\varphi|\leqslant M,\,s\geqslant 0}}(\|Q(s,\ x,\ \varphi)\|+\|f(x)\|)=0,\end{aligned}$$

因此对任一 $\varepsilon>0$，

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\int_{t_k}^{t} r_s(Q(s,\ x_s,\ \varphi_s)-f(x_s))\mathrm{d}s\right\|<\varepsilon,$$

$$\forall t\in[t_k,\ s(t_k,\ T)]$$

最后由定理 5.2.1 便知 $x_t\xrightarrow[t\to\infty]{}x^0$。

第7章

适应性滤波算法

在雷达、通讯等信号处理问题中，在线性系统的参数估计中，常常有这样一类问题：对一个信号列加权，在最小方差意义下去逼近另一个信号列，而这样求得的最优加权阵，或称为*最优线性滤波阵*，涉及矩阵逆或伪逆的计算，在矩阵维数较大或要求实时处理的实际问题中，直接计算的办法不可取或不可能．利用不断得到的这两个信号列的新数据去逐步修正对最优加权阵的估计，为此而构造的递推算法称为*适应性滤波*(adaptive filtering)算法．正如在后面式(7.2.34)、(7.2.35)中给出的算法，我们可以把这类算法变形为线性回归函数的随机逼近算法，但此时的量测误差中含有递推估计值(如(7.2.35))．在实际问题通常所要求的信号条件下，我们将首先证明算法的一致有界性，然后可以验证上述的量测误差满足随机逼近算法的收敛性条件 $A2.4.3$，也即在讨论算法的收敛性问题时，适应性滤波算法实质就是随机逼近算法的特例．另一方面，量测误差中含有递推估计又使得我们在讨论适应性滤波算法的其他渐近性质，如收敛性的充要条件，稳健性，渐近正态性和渐近有效性时不能直接利用前面第2、3、4章的结果，但我们将仍然采用与前几章类似的方法来处理适应性滤波的这些问题．也正因为适应性滤波算法的这个特性，在讨论算法的渐近正态性和渐近有效性这样的随机性质时，与随机逼近算法相比在技术细节上差别更显著些．限于篇幅和为保持全书内容上的一致性，我们在本章只讨论步长因子趋于0的适应性滤波算法，将讨论算法的收敛性，步长因子为$\left\{\frac{a}{n}\right\}$时算法收敛性的充分必要条件及稳健性，

算法的渐近正态性及渐近渐效性，平行地给出与第3、4章内容类似的各种结果. 这一章的主要参考文献为[17, 33～36, 66, 92, 94, 99, 100, 104, 106].

§7.1 最小方差线性滤波

设有两个可观测的随机阵序列$\{y_n\}\in R^{r\times l}$和$\{\psi_n\}\in R^{m\times l}$. 在雷达和通讯问题中，它们可能分别来自适应性阵列处理的输出信号和参考信号；在线性系统的参数估计中，它们又分别为回归阵和系统输出的观测数据. 我们希望求出一个最小方差线性滤波阵x^*，使

$$x^*=\arg\min_x \mathrm{E}(x^\tau y_n-\psi_n)(x^\tau y_n-\psi_n)^\tau. \tag{7.1.1}$$

在实际问题中常常可能对滤波阵有如下约束：

$$x^\tau C=\Phi,\ C\in R^{r\times l},\ \Phi\in R^{m\times l}. \tag{7.1.2}$$

由第一章矩阵伪逆的知识知，式(7.1.2)对x有解的充要条件为$\Phi C^+C=\Phi$. 显然，若取$C=0$, $\Phi=0$, 式(7.1.2)实际成了无约束的情况，也就是说，有约束情况是无约束情况的推广. 因此，本章我们只讨论有约束情况. 在信号处理中，对有约束的适应性滤波通常称为*适应性波束形成*(adaptive beamforming)(参见文献[17]).

在本章中，一直设$\{y_n\}$和$\{\psi_n\}$为平稳序列，并记

$$D_n=(y_n^\tau\psi_n^\tau)^\tau.$$

由平稳性及矩阵对角化方法，

$$\mathrm{E}D_nD_n^\tau=\begin{pmatrix}\mathrm{E}y_ny_n^\tau & \mathrm{E}y_n\psi_n^\tau\\ \mathrm{E}\psi_ny_n^\tau & \mathrm{E}\psi_n\psi_n^\tau\end{pmatrix}=RR^\tau, \tag{7.1.3}$$

其中R可表为分块阵

$$R=\begin{pmatrix}H\\ L\end{pmatrix},$$

使

$$\mathrm{E}y_ny_n^\tau=HH^\tau,\ \mathrm{E}y_n\psi_n^\tau=HL^\tau,\ \mathrm{E}\psi_n\psi_n^\tau=LL^\tau. \tag{7.1.4}$$

下面，利用矩阵伪逆的性质和配方法将给出在约束(7.1.2)下，满足式(7.1.1)的 x^*，即最小方差线性滤波阵.

由式(7.1.1)，$C^\tau x=\Phi^\tau$，再由定理1.4.3知 x 的通解为

$$x=C^{+\tau}\Phi^\tau+(I-C^{\tau+}C^\tau)y, \tag{7.1.5}$$

其中 y 为任意维数相容阵. 注意到 $(CC^+)^\tau=CC^+$，

$$C^{\tau+}C^\tau=C^{+\tau}C^\tau=(CC^+)^\tau=CC^+. \tag{7.1.6}$$

若记

$$P=I-CC^+. \tag{7.1.7}$$

则重写式(7.1.5)为

$$x=C^{+\tau}\Phi^\tau+Py. \tag{7.1.8}$$

于是由式(7.1.1)、(7.1.4)、(7.1.8)和§7.2中式(7.2.3)、(7.2.4)知，

$$\begin{aligned}
&\min_x \mathrm{E}(x^\tau y_n-\psi_n)(x^\tau y_n-\psi_n)^\tau\\
&=\min_y \mathrm{E}[(\Phi C^+ +y^\tau P)y_n-\psi_n][(\Phi C^+ +y^\tau P)y_n-\psi_n]^\tau\\
&=\min_y \mathrm{E}[y^\tau P y_n-(\psi_n-\Phi C^+y_n)][y_n^\tau P y-(\psi_n-\Phi C^+y_n)^\tau]\\
&=\min_y [\mathrm{E}(\psi_n-\Phi C^+y_n)(\psi_n-\Phi C^+y_n)^\tau\\
&\quad -y^\tau P\mathrm{E}y_n(\psi_n-\Phi C^+y_n)^\tau\\
&\quad -\mathrm{E}(\psi_n-\Phi C^+y_n)y_n^\tau Py+y^\tau P\mathrm{E}y_ny_n^\tau Py]\\
&=\min_y [y^\tau-(LH^\tau-\Phi C^+HH^\tau)P(PHH^\tau P)^+]PHH^\tau P\\
&\quad\cdot[y^\tau-(LH^\tau-\Phi C^+HH^\tau)P(PHH^\tau P)^+]^\tau\\
&\quad +\mathrm{E}(\psi_n-\Phi C^+y_n)(\psi_n-\Phi C^+y_n)^\tau-(LH^\tau\\
&\quad -\Phi C^+HH^\tau)P(PHH^\tau P)^+P(LH^\tau-\Phi C^+HH^\tau)^\tau.
\end{aligned} \tag{7.1.9}$$

显然，由式(7.1.9)知

$$y^*=(PHH^\tau P)^+P(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau)+Z. \tag{7.1.10}$$

其中 Z 为任意满足

$$Z^\tau PHH^\tau PZ=0$$

的阵，进而由式(7.1.8)知

$$x^* = C^{+\tau}\Phi^\tau + P(PHH^\tau P)^+ P(HL^\tau - HH^\tau C^{+\tau}\Phi^\tau) + PZ. \tag{7.1.11}$$

当无约束时，即 $C=0$, $\Phi=0$,

$$x^* = (HH^\tau)^+ HL^\tau + Z, \tag{7.1.12}$$

其中 Z 为任意满足

$$Z^\tau HH^\tau Z = 0$$

的阵. 当 HH^τ 非奇异时，x^* 只有唯一解

$$x^* = (HH^\tau)^{-1} HL^\tau, \tag{7.1.13}$$

至此，在式(7.1.11)～(7.1.13)中，给出了几种不同情况下的最小方差线性滤波阵. 在下一节，还将证明

$$P(PHH^\tau P)^+ P = (PHH^\tau P)^+.$$

因而还可重写式(7.1.11)为

$$x^* = C^{+\tau}\Phi^\tau + (PHH^\tau P)^+(HL^\tau - HH^\tau C^{+\tau}\Phi^\tau) + PZ. \tag{7.1.14}$$

注 7.1.1 上述结果还可以推广到一类非平稳的信号情况. 设式(7.1.3)中 R 的定义不变，但有

$$\mathrm{E}D_n D_n^\tau = m_n RR^\tau, \ m_n > 0.$$

因而

$$\mathrm{E}y_n y_n^\tau = m_n HH^\tau, \ \mathrm{E}y_n \psi_n^\tau = m_n HL^\tau, \ \mathrm{E}\psi_n \psi_n^\tau = m_n LL^\tau. \tag{7.1.15}$$

易见，将式(7.1.15)代入(7.1.8)～(7.1.9)的演算中，仍有式(7.1.10)～(7.1.14)成立. 这类情况表明当信号列同步地增强至无穷或衰减至0都仍有不变的最优加权阵，这在实际问题中显然是有意义的. 在[104, 106]中，针对这类非平稳信号分析了算法的强收敛性和收敛速度，但在本书中，为记号上的简单，我们只讨论平稳信号的情况，然而要推广本书的结果至式(7.1.15)并无本质困难.

注 7.1.2 一般来说，式(7.1.14)给出了由 x^* 组成的集合，但在实际工程中，人们常常假设 HH^τ 非奇异，由式(7.1.10)后 z 的定义可得 $PZ=0$，于是，我们欲求的最小方差线性滤波阵为

$$x^* = C^{+\tau}\Phi^\tau + (PHH^\tau P)^+(HL^\tau - HH^\tau C^{+\tau}\Phi^\tau). \tag{7.1.16}$$

§7.2 适应性滤波算法的强收敛性

在上节，已给出了最小方差线性滤波阵的精确表达式(7.1.16)，但由于 HH^τ 和 HL^τ 一般并不知道，直接计算 x^* 是不可能的. 即使可以利用信号的平稳性来估计这两个阵，当信号的维数较大时，计算矩阵逆或伪逆仍很慢，不利于实时处理. 于是，在实际问题中，人们不采用这种直接估算的办法，而是利用不断得到的信号去递推地估计 x^*，考虑如下的适应性滤波算法：

$$x_{n+1} = C^{+\tau}\Phi^\tau + P[x_n + a_n(y_n\psi_n^\tau - y_n y_n^\tau x_n)], \quad x_0 = C^{+\tau}\Phi^\tau, \tag{7.2.1}$$

其中 $\{a_n\}$ 为正实数序列的步长因子.

为了便于分析算法(7.2.1)，我们还需要对它作一些变形.

不失一般性，可设 $PH \neq 0$，否则从式(7.1.14)知 $x^* = C^{+\tau}\Phi^\tau$，不需要任何算法. 记 $L(A)$ 为阵 A 列向量张成的向量子空间，设 V_1 的列向量为 $L(C)$ 的直交规范基，V_2 的列向量为 $L(PH)$ 的直交规范基. 由于从

$$C^\tau P = C^\tau - C^\tau CC^+ = 0,$$

知 $L(C) \perp L(PH)$，也即

$$V_1^\tau V_2 = 0 \quad (\text{不一定为方阵}). \tag{7.2.2}$$

显然存在 V_3，使得 $[V_1 V_2 V_3]$ 为直交阵.

由 V_2 的定义知，存在 PH 的满秩分解如下

$$PH = V_2 K, \quad KK^\tau > 0, \tag{7.2.3}$$

进而由伪逆性质(定理 1.4.2, 2))

$$PHH^\tau P = V_2 KK^\tau V_2^\tau, \quad (PHH^\tau P)^+ = V_2(KK^\tau)^{-1}V_2^\tau. \tag{7.2.4}$$

于是由 $L(C) \perp L(PH)$、V_2 定义和式(7.1.6)、(7.2.4)，得

$$P(PHH^\tau P)^+ P = (PHH^\tau P)^+, \tag{7.2.5}$$

此即式(7.1.11)改写为式(7.1.14)的根据.

类似,C 也有满秩分解,

$$C=V_1K_1,\ K_1K_1^\tau>0, \tag{7.2.6}$$

于是由§1.4矩阵伪逆的表达式(1.4.6),

$$\begin{aligned} C^+&=K_1^\tau(K_1K_1^\tau)^{-1}(V_1^\tau V_1)^{-1}V_1^\tau, \\ CC^+&=V_1V_1^\tau. \end{aligned} \tag{7.2.7}$$

由式(7.2.7),可写

$$\begin{aligned} P&=I-CC^+=[V_1V_2V_3][V_1V_2V_3]^\tau-CC^+ \\ &=V_1V_1^\tau+V_2V_2^\tau+V_3V_3^\tau-CC^+=V_2V_2^\tau+V_3V_3^\tau. \end{aligned} \tag{7.2.8}$$

注意到

$$V_3^\tau H=V_3^\tau(P+CC^+)H=V_3^\tau PH+V_3^\tau CC^+H=0,$$

故由式(7.2.8)、(7.2.4),得

$$PH=V_2V_2^\tau H,\ PV_2=V_2,\ V_2^\tau P=V_2^\tau \tag{7.2.9}$$

$$PHH^\tau P(PHH^\tau P)^+H=PH. \tag{7.2.10}$$

以上矩阵等式将在以后用到.

由式(7.1.8)和(7.2.1),对任意 $n\geqslant 0$, $x_n^\tau C=\Phi$,于是设

$$S_n=Px_n,\ S^*=Px^*, \tag{7.2.11}$$

可得

$$S_n^\tau=x_n^\tau(I-CC^+)=x_n^\tau-\Phi C^+,\ S^*=x^*-C^{+\tau}\Phi^\tau. \tag{7.2.12}$$

由式(7.2.1)、(7.2.11)和(7.2.12)得

$$\begin{aligned} S_{n+1}&=S_n+a_nP(y_n\psi_n^\tau-y_ny_n^\tau(S_n+C^{+\tau}\Phi^\tau)), \\ S_0&=0. \end{aligned} \tag{7.2.13}$$

由于

$$\mathrm{E}V_3^\tau Py_ny_n^\tau PV_3=0,\ \mathrm{E}V_1^\tau Py_ny_n^\tau PV_1=0,$$

故

$$V_3^\tau Py_n=0,\ \text{a.s.},\ V_1^\tau Py_n=0,\quad \text{a.s.},$$

即 $Py_n\in L(PH)$,再由式(7.2.13)知

$$S_n\in L(PH),\quad \text{a.s.}.$$

因而对任 $n\geqslant 0$，存在唯一的阵 $Q_n\in R^{d\cdot mL(PH)\times r}$，使得 $S_n=V_2Q_n$，这意味着式(7.2.13)可改写为

$$V_2Q_{n+1}=V_2Q_n+a_nP[y_n\psi_n^\tau-y_ny_n^\tau(V_2Q_n+C^{+\tau}\Phi^\tau)].\tag{7.2.14}$$

注意到式(7.2.9)中的 $PV_2=V_2$，因而有

$$\begin{aligned}Q_{n+1}&=Q_n+a_nV_2^\tau P[y_n\psi_n^\tau-y_ny_n^\tau(V_2Q_n+C^{+\tau}\Phi^\tau)]\\&=(I-a_nV_2^\tau Py_ny_n^\tau PV_2)Q_n\\&\quad+a_nV_2^\tau P(y_n\psi_n^\tau-y_ny_n^\tau C^{+\tau}\Phi^\tau)\\&=(I-a_nz_nz_n^\tau)Q_n+a_nG_n,\ Q_0=0,\end{aligned}\tag{7.2.15}$$

其中

$$z_n=V_2^\tau Py_n,\ G_n=V_2^\tau P(y_n\psi_n^\tau-y_ny_n^\tau C^{+\tau}\Phi^\tau),\tag{7.2.16}$$

且由式(7.2.3)易见

$$\mathrm{E}z_nz_n^\tau=KK^\tau>0.\tag{7.2.17}$$

由式(7.1.12)知

$$x_n=C^{+\tau}\Phi^\tau+S_n=C^{+\tau}\Phi^\tau+V_2Q_n,\tag{7.2.18}$$

再由式(7.2.4)得

$$V_2V_2^\tau(PHH^\tau P)^+=(PHH^\tau P)^+.\tag{7.2.19}$$

因此，由式(7.1.16)，(7.2.18)和(7.2.19)可得对应于 x^* 的 Q^* 为

$$Q^*=V_2^\tau(PHH^\tau P)^+(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau).\tag{7.2.20}$$

至此，我们将算法(7.2.1)中的 x_n 收敛于 x^* 的问题化为了算法(7.2.15)中的 Q_n 收敛于 Q^* 的问题.

为讨论算法的强收敛性，需要下列关于步长因子$\{a_n\}$和信号$\{y_n\}$和$\{\psi_n\}$的假设 A7.2:

A 7.2.1 $\qquad a_n>0,\ a_n\xrightarrow[n]{}0,\ \sum_{n=0}^{\infty}a_n=\infty.$

A 7.2.2

$$\sum_{n=0}^{\infty}a_n(y_ny_n^\tau-HH^\tau)\quad 收敛\ \text{a.s.},$$

$$\sum_{n=0}^{\infty} a_n (y_n \psi_n^\tau - HL^\tau) \quad 收敛\ \text{a.s.}.$$

注 7.2.1 正如在§2.3中所说，满足这里假设的信号可以是无穷相关的，包括了工程中常用到各种随机序列.

下面，我们先证明几个引理，然后再证明算法(7.2.15)的有界性，最后用§2.4中 Lyapunov 函数方法建立算法强收敛性的定理(参见文献[106]).

定义

$$U_{nk} = \begin{cases} \prod_{i=k+1}^{n} (I - a_i z_i z_i^\tau), & n > k, \\ I, & n = k. \end{cases} \tag{7.2.21}$$

对任 $m \geqslant n \geqslant 0$，重写式(7.2.15)

$$\begin{aligned} Q_{m+1} - Q^* &= (I - a_m z_m z_m^\tau)(Q_m - Q^*) + a_m (G_m - z_m z_m^\tau Q^*) \\ &= U_{m\,n-1} (Q_n - Q^*) + \sum_{j=n}^{m} a_j U_{mj} (G_j - z_j z_j^\tau Q^*). \end{aligned} \tag{7.2.22}$$

引理 7.2.1 在假设 A7.2 下，

$$\sum_{n=0}^{\infty} a_n (G_n - z_n z_n^\tau Q^*) \quad 收敛\ \text{a.s.}.$$

证明 利用式(7.2.10)、(7.2.16)、(7.2.19)和(7.2.20)，可以重写

$$\begin{aligned} &\sum_{n=0}^{\infty} a_n (G_n - z_n z_n^\tau Q^*) \\ &= \sum_{n=0}^{\infty} a_n V_2^\tau [P(y_n \psi_n^\tau - y_n y_n^\tau C^{+\tau} \Phi^\tau) \\ &\quad - P y_n y_n^\tau P (PHH^\tau P)^+ (HL^\tau - HH^\tau C^{+\tau} \Phi^\tau) \\ &\quad + PHH^\tau P (PHH^\tau P)^+ (HL^\tau - HH^\tau C^{+\tau} \Phi^\tau) \\ &\quad - P(HL^\tau - HH^\tau C^{+\tau} \Phi^\tau)]. \end{aligned}$$

从上式和 A7.2 立即可知引理成立. □

由式(7.2.17)知 $KK^\tau = V^\tau PHH^\tau PV$ 是正定阵. 设 λ 是 KK^τ 的最小本征值.

引理 7.2.2 在假设 A7.2 下，存在 $\eta > 0$, $\eta\lambda < 1$，和 K_η，使得

$$\left\|\prod_{j=n}^{m(n,\eta)-1}(I-a_jz_jz_j^{\tau})\right\|<1-\frac{1}{2}\eta\lambda,\ \forall n>K_\eta,\quad \text{a.s.},$$

其中 $$m(n,\ \eta)=\max\{m:\sum_{j=n}^{m}a_j\leqslant\eta\},$$

矩阵模 $$\|A\|=\max[\lambda(AA^{\tau})]^{\frac{1}{2}}.$$

证明 由于本引理为a.s.成立的命题，因此以下讨论均对固定的ω而言.

对任意序列$\{\mu_i\}$定义

$$H_n^m(\mu_i)=\sum_{d=2}^{m-n+1}\sum_{n\leqslant i_1<\cdots<i_d\leqslant m}\prod_{j=1}^{d}\mu_{i_j},$$

则

$$\prod_{j=n}^{m(n,\eta)-1}(I-a_jz_jz_j^{\tau})=I-\sum_{d=n}^{m(n,\eta)-1}a_jz_jz_j^{\tau}+H_n^{m(n,\eta)-1}(-a_iz_iz_i^{\tau}),\tag{7.2.23}$$

$$\prod_{j=n}^{m(n,\eta)-1}(1+a_j\|z_j\|^2)=1+\sum_{j=n}^{m(n,\eta)-1}a_j\|z_j\|^2+H_n^{m(n,\eta)-1}\cdot(a_i\|z_i\|^2).\tag{7.2.24}$$

另一方面，设$z=\sum_{j=n}^{m(n,\eta)-1}a_j\|z_j\|^2/(m(n,\ \eta)-n)$，则

$$\begin{aligned}\prod_{j=n}^{m(n,\eta)-1}(1+a_j\|z_j\|^2)&\leqslant\left(\frac{\sum_{j=n}^{m(n,\eta)-1}(1+a_j\|z_j\|^2)}{m(n,\ \eta)-n}\right)^{m(n,\eta)-n}\\&=(1+z)^{m(n,\eta)-n}=1+(m(n,\ \eta)-n)z\\&\quad+H_n^{m(n,\eta)-1}(z)\\&=1+\sum_{j=n}^{m(n,\eta)-1}a_j\|z_j\|^2+H_n^{m(n,\eta)-1}(z),\end{aligned}\tag{7.2.25}$$

其中 $$H_n^{m(n,\eta)-1}(z)=\sum_{d=2}^{m(n,\eta)-n}\sum_{n\leqslant i_1<\cdots<i_d\leqslant m(n,\eta)-1}z^d.$$

比较式(7.2.24)和(7.2.25)，我们得到

$$\begin{aligned}\|H_n^{m(n,\eta)-1}(-a_iz_iz_i^{\tau})\|&\leqslant H_n^{m(n,\eta)-1}(a_i\|z_i\|^2)\\&\leqslant H_n^{m(n,\eta)-1}(z).\end{aligned}\tag{7.2.26}$$

由 A7.2.2，并记 $\|A\|_* = (\operatorname{tr} AA^\tau)^{\frac{1}{2}}$，可得

$$\sum_{j=n}^{m(n,\eta)-1} a_j \|z_j\|_*^2 \xrightarrow[n]{} \eta \|K\|_*^2, \tag{7.2.27}$$

其中 $K = V_2 PH$. 再注意

$$\|A\| = \max(\lambda(AA^\tau))^{\frac{1}{2}} \leqslant \|A\|_*,$$

$$H_n^{m(n,\eta)-1}(z) \leqslant H_n^{m(n,\eta)-1}\left(\sum_{j=n}^{m(n,\eta)-1} a_j \|z_j\|_*^2 / (m(n,\eta)-n)\right)$$

$$= \left(1 + \frac{\sum_{j=n}^{m(n,\eta)-1} a_j \|z_j\|_*^2}{m(n,\eta)-n}\right)^{m(n,\eta)-n}$$

$$- \left(1 + \sum_{j=n}^{m(n,\eta)-1} a_j \|z_j\|_*^2\right) \xrightarrow[n]{} e^{\eta\|K\|_*^2} - (1 + \eta\|K\|_*^2)$$

$$= \sum_{j=2}^{\infty} \frac{(\eta\|K\|_*)^2}{j!} \leqslant \eta^2 \|K\|_*^4 e^{\eta\|K\|_*^2}, \tag{7.2.28}$$

显然，上式右边当 $\eta \to 0$，为 $o(\eta)$ 量. 因此存在 η 和正实数 k_η，使

$$H_n^{m(n,\eta)}(z) \leqslant \frac{1}{3}\eta\lambda, \quad \forall n \geqslant k_\eta, \tag{7.2.29}$$

由 A7.2.2，我们还有

$$\left\|\sum_{j=n}^{m(n,\eta)-1} a_j z_j z_j^\tau - \eta K K^\tau\right\| < \frac{1}{6}\eta\lambda, \quad \forall n \geqslant k_\eta. \tag{7.2.30}$$

于是对任意 $n \geqslant k_\eta$，由式(7.2.23)、(7.2.26)、(7.2.29)和(7.2.30)得

$$\left\|\prod_{j=n}^{m(n,\eta)-1} (I - a_j z_j z_j^\tau)\right\| \leqslant \left\|I - \sum_{j=n}^{m(n,\eta)-1} a_j z_j z_j^\tau\right\|$$

$$+ \|H_n^{m(n,\eta)-1}(-a_i z_i z_i^\tau)\|$$

$$\leqslant \|I - \eta K K^\tau\| + \left\|\sum_{j=n}^{m(n,\eta)-1} a_j z_j z_j^\tau - \eta k k^\tau\right\|$$

$$+ H_n^{m(n,\eta)-1}(z) \leqslant 1 - \frac{1}{2}\eta\lambda.$$

□

引理 7.2.3　在假设 A7.2 下，存在 $\eta > 0$ 和 k_η，使对任意 $n \geqslant k_\eta$ 和 $m \geqslant m(n, \eta)$，

$$\left\|\prod_{j=n}^{m} (I - a_j z_j z_j^\tau)\right\| \leqslant \exp\left(-\frac{\lambda}{2}\sum_{j=n}^{m} a_j\right).$$

证明 从引理 7.2.2, 存在 $\eta>0$ 和 k_η, 对任意 $n\geqslant k_\eta$, 有

$$\left\|\prod_{j=n}^{m(n,\eta)-1}(I-a_j z_j z_j^\tau)\right\|\leqslant 1-\frac{1}{2}\eta\lambda\leqslant e^{-\frac{1}{2}\eta\lambda}. \qquad (7.2.31)$$

对任意 $m\geqslant m(n,\ \eta)$ 和 $n\geqslant k_\eta$, 定义 $n_0<n_1<\cdots<n_r$, 使

$$n=n_0<m(n_0,\ \eta)=n_1<\cdots<m(n_{r-2},\ \eta)$$

$$=n_{r-1}<n_r=m,\ 且\ \sum_{i=n_{r-1}}^{n_r}a_i\leqslant\eta.$$

注意到 $\sum_{i=n}^{m}a_i=r\eta$, 由式(7.2.31)

$$\left\|\prod_{j=n}^{m}(I-a_j z_j z_j^\tau)\right\|\leqslant\prod_{i=1}^{r}\left\|\prod_{j=n_{i-1}}^{n_i-1}(I-a_j z_j z_j^\tau)\right\|$$

$$\leqslant\exp\left(-\frac{\lambda}{2}\sum_{j=n}^{m}a_j\right). \qquad \square$$

有了引理 7.2.1 和 7.2.3, 我们可以得到算法的有界性.

定理 7.2.1 当假设 A 7.2 成立, 由式(7.2.15)给出的 Q_n a.s. 一致有界.

证明 我们从式(7.2.22)出发. 据引理 7.2.3, 对任意 $n\geqslant k_\eta$, 式(7.2.22)右端第一项随 $m\to\infty$ 而趋于 0.

记

$$m'=\min\left\{k:\ \sum_{j=k+1}^{m}a_j\leqslant\eta\right\},\ \eta\ 定义如引理\ 7.2.2,$$

$$W_n^m=\sum_{j=n}^{m}a_j(G_j-z_j z_j^\tau Q^*).$$

据引理 7.2.1 和 7.2.3, 并注意式(7.2.16)和(7.2.27)的推导, 用分部求和法知, 式(7.2.22)右端第二项可写成

$$\left\|\sum_{j=n}^{m'}(U_{mj}-U_{m(j+1)})W_n^j+U_{m(m'+1)}W_n^{m'}\right.$$

$$\left.+\sum_{j=m'+1}^{m}a_j U_{mj}(G_j-z_j z_j^\tau Q^*)\right\|$$

$$\leqslant\sum_{j=n}^{m'}\|U_{m(j+1)}\|\,\|a_j z_j z_j^\tau\|\,\|W_n^j\|+\|U_{m(m'+1)}\|\,\|W_n^{m'}\|$$

$$+\sum_{j=m'+1}^{m}a_j\|U_{mj}\|(\|G_j\|+\|z_j z_j^\tau Q^*\|)$$

$$\leqslant c\left[\sum_{j=\lambda}^{m'}\exp\left[-(\lambda/2)\sum_{i=j}^{m}a_i\right]\|a_jz_jz_j^{\tau}\|+1\right],\tag{7.2.32}$$

其中 c 及以下的 c_i 为正常数.

由 A 7.2.2 易得, 当 $n\geqslant N$, 对任 $m\geqslant n$,

$$\sum_{j=n}^{m}\|a_jz_jz_j^{\tau}\|\leqslant\sum_{j=n}^{m}a_j\|z_j\|^2\leqslant\|K\|^2\sum_{j=n}^{m}a_j+c_1,$$

故

$$\sum_{j=n}^{m}a_j\geqslant c_2\sum_{j=n}^{m}\|a_jz_jz_j^{\tau}\|-c_1.$$

于是

$$\begin{aligned}&\sum_{j=n}^{m'}\exp\left[-(\lambda/2)\sum_{i=j}^{m}a_i\right]\|a_jz_jz_j^{\tau}\|\\&\leqslant\sum_{j=n}^{m'}\exp\left[-(c_2\eta\lambda/2)\sum_{i=j}^{m}\|a_iz_iz_i^{\tau}\|+c_4\right]\|a_jz_jz_j^{\tau}\|\\&\leqslant c_5\sum_{j=n}^{m}\exp\left[-(c_2\eta\lambda/2)\sum_{i=j}^{m}\|a_iz_iz_i^{\tau}\|\right]\|a_jz_jz_j^{\tau}\|\\&\leqslant c_6\int_0^{\infty}e^{-c_2\eta\lambda x/2}\mathrm{d}x\leqslant c_7.\end{aligned}\tag{7.2.33}$$

于是从式(7.2.32)和(7.2.33)知算法有界. □

下面, 我们用 Lyapunov 函数方法来证明算法的 a.s. 收敛性.

定理 7.2.2 在假设 A7.2 之下, 算法(7.2.1)和(7.2.15)分别 a.s. 收敛到相应的最优值 x^* 和 Q^*.

证明 我们仅证明式(7.2.15)的收敛性. 从式(7.2.15)知

$$Q_{n+1}=Q_n-a_nKK^{\tau}(Q_n-Q^*)+a_n\varepsilon_{n+1},\tag{7.2.34}$$

$$\varepsilon_{n+1}=(KK^{\tau}-z_nz_n^{\tau})(Q_n-Q^*)+G_n-z_nz_n^{\tau}Q^*.\tag{7.2.35}$$

这样, 算法(7.2.15)经过上述变形成了线性回归函数的随机逼近算法, 为证明算法的收敛性, 用 §2.4 的已知结果, 只需验证对算法$\{Q_n\}$, 条件 A2.4.3 成立, 也即设子列 $Q_{n_k}\xrightarrow[k\to\infty]{}Q$, 要验证

$$\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\sum_{i=n_k}^{m(n_k,t)}a_i\varepsilon_{i+1}=0,\ \forall t\in[0,T].\tag{7.2.36}$$

用式(7.2.35)和引理7.2.1,只要验证

$$\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\sum_{i=n_k}^{m(n_k,t)}a_i(KK^\tau-z_iz_i^\tau)(Q_i-Q)=0. \tag{7.2.37}$$

从A7.2.2知

$$\sum_{i=n_k}^{m(n_k,t)}a_i(z_iz_i^\tau-KK^\tau)\xrightarrow[k\to\infty]{}0,$$

或

$$\sum_{i=n_k}^{m(n_k,t)}a_i\|z_i\|_*^2\xrightarrow[k\to\infty]{}t\cdot\operatorname{tr}KK^\tau. \tag{7.2.38}$$

又由定理7.2.1知Q_n一致有界，并注意Q_{n_k}的极限为Q，从式(7.2.34)、(7.2.35)便知当$k\to\infty$时，对一切m: $n_k\leqslant m\leqslant m(n_k,t)$有

$$\|Q_m-Q\|_*=O(t). \tag{7.2.39}$$

由此及式(7.2.38)可得

$$\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\left\|\sum_{i=n_k}^{m(n_k,t)}a_i(KK^\tau-z_iz_i^\tau)(Q_i-Q)\right\|_*$$

$$\leqslant\lim_{T\to 0}\lim_{k\to\infty}\frac{1}{T}\sum_{i=n_k}^{m(n_k,t)}a_i(\|K\|_*^2+\|z_i\|_*^2)O(t)$$

$$=\lim_{T\to 0}\frac{1}{T}(t\|K\|_*^2+t\operatorname{tr}KK^\tau)O(t)=0,\ \forall t\in[0,T].$$

此即式(7.2.37)成立. 所以用定理2.4.1,知

$$Q_n\xrightarrow[n\to\infty]{}Q^*\quad\text{a.s.}. \qquad \square$$

现在,我们进一步考虑算法的强收敛速度，为此，我们需要更多一些条件:

A7.2.3 $a_{n+1}^{-1}-a_n^{-1}\xrightarrow[n]{}\alpha\geqslant 0$.

A7.2.4 存在正数$\delta<1$, $\delta<\dfrac{\lambda}{\alpha}$,使得

$$\sum_{n=1}^{\infty}a_n^{1-\delta}(y_ny_n^\tau-HH^\tau)\quad\text{a.s. 收敛},$$

$$\sum_{n=1}^{\infty}a_n^{1-\delta}(y_n\psi_n^\tau-HL^\tau)\quad\text{a.s. 收敛}.$$

定理7.2.3 在假设A7.2.1~A7.2.4下，算法(7.2.1)强收

敛速度为

$$\|x_n - x^*\| = o(a_n^\delta).$$

证明 仍只需证明 $\|Q_n - Q^*\| = o(a_n^\delta)$, 为此仍从式(7.2.22)出发. 为书写简单, 记

$$X_n = Q_n - Q^*,\ \varepsilon_n = G_n - z_n z_n^\tau Q^*. \tag{7.2.40}$$

再由 $(a_n/a_{n+1})^\delta$ 的 Taylor 展式, 问题化为讨论

$$\begin{aligned}\frac{X_{n+1}}{a_{n+1}^\delta} &= \left(1 + \delta\,\frac{a_n - a_{n+1}}{a_{n+1}} + O\left[\left(\frac{a_n - a_{n+1}}{a_{n+1}}\right)^2\right]\right) \\ &\quad \cdot \left(\frac{X_n}{a_n^\delta} - a_n z_n z_n^\tau \frac{X_n}{a_n^\delta} + a_n^{1-\delta}\varepsilon_n\right) \\ &= \frac{X_n}{a_n^\delta} - a_n(z_n z_n^\tau - \alpha\delta I + e_n I)\,\frac{X_n}{a_n^\delta} + a_n^{1-\delta}\varepsilon_n + a_n\xi_n,\end{aligned} \tag{7.2.41}$$

是否收敛于 0, 其中 e_n, ε_n 和 ξ_n 由 A7.2.3, A7.2.4 和引理 7.2.1 易知它们满足

$$e_n \xrightarrow[n]{} 0 \text{ a.s.},\ \sum_{n=1}^{\infty} a_n^{1-\delta}\varepsilon_n \text{ 收敛},\ \xi_n \xrightarrow[n]{} 0 \ \text{ a.s.}. \tag{7.2.42}$$

记

$$E_n E_n^\tau = z_n z_n^\tau - (\alpha\delta - e_n) I,\ EE^\tau = KK^\tau - \alpha\delta I > 0. \tag{7.2.43}$$

用

$$U'_{nk} = \begin{cases} \prod\limits_{i=k+1}^{n} (I - a_i E_n E_n^\tau), & n > k, \\ I, & n = k \end{cases} \tag{7.2.44}$$

取代式(7.2.21)的 U_{nk}, 注意 $e_n \xrightarrow[n\to\infty]{} 0$, 和引理 7.2.2 类似地可证, 对充分大的 n

$$\|U_{m(n,\eta)-1,n-1}\| < 1 - \frac{1}{2}\eta\lambda', \tag{7.2.45}$$

其中 λ' 为 EE^τ 的最小本征值. 这样, 照样可证和引理 7.2.3 及定理 7.2.1 类似的结论, 因此 X_{n+1}/a_{n+1}^δ 有界.

记

$$Q_n' = X_n/a_n^\delta,\ \varepsilon_{n+1}' = (EE^\tau - z_n z_n^\tau)Q_n' + \varepsilon_n/a_n^\delta + \xi_n.$$

完全和定理 7.2.2 类似地可证 $Q_n' \xrightarrow[n\to\infty]{} 0$ a.s.. □

为了使读者对算法的收敛速度有一个具体的印象，我们在下面的推论中给出一个例子.

推论 7.2.1 设 $a_n = \frac{1}{n}$,信号是平稳的,且

$$y_n y_n^\tau = HH^\tau + w_n + B_1 w_{n-1} + \cdots + B_k w_{n-k} + \cdots,$$

$$y_n \psi_n^\tau = HL^\tau + u_n + C_1 u_{n-1} + \cdots + C_k u_{n-k} + \cdots,$$

其中$\|B_n\| \leqslant c/n^\gamma$，$\gamma > \frac{3}{2}$；$\|C_n\| \leqslant c/n^\beta$，$\beta > \frac{3}{2}$，$w_n$ 和 u_n 均为鞅差序列,满足

$$\mathrm{E}(w_{n+1}/\mathscr{F}_n) = \mathrm{E}(u_{n+1}/\mathscr{F}_n) = 0,$$

$$\mathrm{E}(\|w_{n+1}\|^2/\mathscr{F}_n) < c,\ \mathrm{E}(\|u_{nn}\|^2/\mathscr{F}_n) < c.$$

此外，还设引理 7.2.2 中的$\lambda \geqslant \frac{1}{2}$,则有

$$\|x_n - x^*\| = o\left(\frac{1}{n^\delta}\right),\ \forall \delta < \frac{1}{2}.$$

证明 由对信号的假设，$\{y_n y_n^\tau - HH^\tau\}$和$\{y_n \psi_n^\tau - HL^\tau\}$均满足例 2.3.1 中的条件,因此

$$\sum_{i=1}^{\infty} \frac{1}{n^{1-\delta}}(y_n y_n^\tau - HH^\tau) \quad \text{a.s. 收敛},\ \forall 0 \leqslant \delta < \frac{1}{2},$$

$$\sum_{i=1}^{\infty} \frac{1}{n^{1-\delta}}(y_n y_n^\tau - HH^\tau) \quad \text{a.s. 收敛},\ \forall 0 \leqslant \delta < \frac{1}{2}.$$

显然对$a_n = \frac{1}{n}$，A7.2.1 和 A7.2.3 也成立,且此时$\alpha = 1$,因而$\lambda > \alpha\delta = \delta$. 于是由定理 7.2.3,本推论成立. □

由这个例子可以看出，这类适应性滤波算法的收敛速度通常不超过$\frac{1}{\sqrt{n}}$，与随机逼近算法类似. 所以，虽然这类步长因子衰减到 0 的适应性滤波算法具有算法简单,大样本性质好的优点,但也有收敛速度较慢的局限,在某些需要高速实时处理的问题中，人

们需要考虑别的算法，如步长因子为一个小常值的算法，有兴趣的读者可参考文献[46, 55, 92].

§7.3 一类步长因子下强收敛性的充要条件及稳健性

上节分析了一般的步长因子(定义如 A7.2.1)的适应性滤波算法的强收敛性，这一节我们对一类特殊的步长因子，$a_n=\frac{a}{n}$，$a>0$，讨论算法强收敛的充分必要条件．我们采用不同于上节的先证算法有界性，再证算法收敛性的方法，而是直接分析获得算法的收敛性．最后得到相当简单的充分必要条件；粗略地说，当$\{y_ny_n^\tau\}$满足大数定律，则适应性滤波算法强收敛的充分必要条件是$\{y_n\psi_n^\tau\}$也满足大数定律．

此外，我们还分析了这类算法强收敛的稳健性问题．结果是：当对$\{y_ny_n^\tau\}$和$\{y_n\psi_n^\tau\}$大数定律并不严格成立，而存在小偏差时，算法对真值也存在一个依赖于前者的偏差上界，当前者趋于 0，后者也趋于 0．

下面，先给出本节需要的假设：

A7.3.1 $\lim\limits_{n}\frac{1}{n}P\sum\limits_{j=1}^{n}y_jy_j^\tau P=PHH^\tau P$, a.s.,

A7.3.2 $\lim\limits_{n}\frac{1}{n}P\sum\limits_{j=1}^{n}y_j\psi_j^\tau=PHL^\tau$, a.s..

显然，A7.3.1 和 A7.3.2 较 A7.2.2 更一般，因为若 7.2.2 成立，则根据引理 1.2.6(Kronecker 引理)，必有 A7.3.1 和 A7.3.2 成立．

进一步，如果$\{y_ny_n^\tau\}$和$\{y_n\psi_n^\tau\}$并不严格满足上述大数定律，而有如下偏差：

A7.3.3 $\frac{1}{n}P\sum\limits_{j=1}^{n}y_jy_j^\tau P=PHH^\tau P+\varepsilon_n$, $\|\varepsilon_n\|<\varepsilon$, a.s. $n\geqslant N$,

A7.3.4 $\frac{1}{n}P\sum\limits_{j=1}^{n}y_j\psi_j^\tau=PHL^\tau+\eta_n$, $\|\eta_n\|<\varepsilon$, a.s. $n\geqslant N$,

其中 s 是某正常数，N 可能依赖于抽样点 ω，将讨论算法的稳健性.

本节所讨论的算法为：

$$x_{n+1}=C^{+\tau}\Phi^{\tau}+P\left[x_n+\frac{a}{n}\left(y_n\psi_n^{\tau}-y_ny_n^{\tau}x_n\right)\right],\ x_0=C^{+\tau}\Phi^{\tau}. \tag{7.3.1}$$

和上节一样，仍从(7.3.1)的如下变形

$$Q_{m+1}-Q^{*}=U_{mn}(Q_n-Q^{*})+\sum_{j=n}^{m}\frac{a}{j}U_{mj}(G_j-z_jz_j^{\tau}Q^{*}),\ Q_0=0 \tag{7.3.2}$$

出发，其中 z_n, G_n 定义如式(7.2.16)，U_{mj} 定义如式(7.2.21)．只是其中的 $a_i=\frac{a}{i}$.

正如定理 7.2.2 证明过程中所表明，算法(7.3.2)可化为由式(7.2.34)和(7.2.35)所描述的随机逼近算法，类似于定理 3.4.1，我们有如下定理.

定理 7.3.1 假设 A7.3.1 成立，则算法(7.3.1)a.s. 收敛到 x^{*} 的充分必要条件为 A7.3.2 成立.

证明 仍从算法(7.3.2)出发

(i) 充分性 当 A7.3.1 成立，先证明下面两个引理.

引理 7.3.1 在 A7.3.1 之下，存在 $\eta>0$, $\eta\lambda<1$ 和 K_η，使得对任意 $n\geqslant K_\eta$, $m\geqslant m(n,\eta)$,

$$\left\|\prod_{j=n}^{m}\left(I-\frac{a}{j}z_jz_j^{\tau}\right)\right\|\leqslant\exp\left(-\frac{\lambda}{2}\sum_{j=n}^{m}a/j\right).$$

证明 类似于上一节的引理 7.2.3 的证明，我们先证明在 A7.3.1 之下，存在 $\eta>0$, $\eta\lambda<1$ 和 k_η，使得对任意 $n\geqslant k_\eta$,

$$\prod_{j=n}^{m(n,\eta)-1}\left(I-\frac{a}{j}z_jz_j^{\tau}\right)<1-\eta\lambda/2. \tag{7.3.3}$$

类似于式(7.2.23)，这里仍有

$$\prod_{j=n}^{m(n,\eta)-1}\left(I-\frac{a}{j}z_jz_j^{\tau}\right)=I-\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}z_jz_j^{\tau}+H_n^{m(n,\eta)-1}\left(-\frac{a}{j}z_jz_j^{\tau}\right). \tag{7.3.4}$$

记

$$W_i=\left(\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}\|z_j\|^{\tau}\right)^i. \tag{7.3.5}$$

显然，在上式右端的展开式中若只取 $\frac{a}{j}\|z_j\|^2$ 相异乘积的项，则

$$W_i\geqslant i!\sum_{n<j_1<\cdots<j_i\leqslant m(n,\eta)-1}\frac{a^i\|z_{j_1}\|^2\cdots\|z_{j_i}\|^2}{j_1j_2\cdots j_i}. \tag{7.3.6}$$

于是由式(7.2.26)和(7.3.6)，我们有

$$\left\|H_n^{m(n,\eta)-1}\left(-\frac{a}{j}z_jz_j^{\tau}\right)\right\|\leqslant H_n^{m(n,\eta)-1}\left(\frac{a}{j}\|z_j\|^2\right)\leqslant\sum_{j=2}^{m(n,\eta)-n}\frac{W_j}{j!}. \tag{7.3.7}$$

记

$$S_n=\sum_{j=1}^{n}z_jz_j^{\tau}. \tag{7.3.8}$$

注意到$\|A\|^2\leqslant\operatorname{tr}AA^{\tau}$，并应用分部求和法，

$$\begin{aligned}\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}\|z_j\|^2&\leqslant\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}\operatorname{tr}(S_j-S_{j-1})\\&=\operatorname{tr}\sum_{j=n-1}^{m(n,\eta)-2}\frac{a}{j(j+1)}S_j\\&\quad+\frac{a}{m(n,\eta)-1}\operatorname{tr}S_{m(n,\eta)-1}-\frac{a}{n-1}\operatorname{tr}S_{n-1}.\end{aligned} \tag{7.3.9}$$

由 A7.3.1 知，$\lim\limits_{n}S_n/n=KK^{\tau}$，因而对任 $\delta>0$，定存在 n_0，使当 $n\geqslant n_0$，

$$\left|a\operatorname{tr}\left(\frac{S_n}{n}-KK^{\tau}\right)\right|\leqslant\delta,\quad\left\|a\left(\frac{S_n}{n}-KK^{\tau}\right)\right\|\leqslant\delta,$$

$$\frac{a}{n}\|KK^{\tau}\|\leqslant\frac{1}{2}\delta. \tag{7.3.10}$$

从式(7.3.9)和(7.3.10)，并注意$\|A\|^2\leqslant\operatorname{tr}AA^{\tau}$，

$$\begin{aligned}\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}\|z_j\|^2&\leqslant(a\operatorname{tr}KK^{\tau}+\delta)\sum_{j=n-1}^{m(n,\eta)-2}\frac{1}{j+1}+2\delta\\&\leqslant\eta(\operatorname{tr}KK^{\tau}+\delta)+2\delta.\end{aligned} \tag{7.3.11}$$

故从式(7.3.5)、(7.3.7)和(7.3.11)得

$$\left\|H_n^{m(n,\eta)-1}\left(-\frac{a}{j}z_jz_j^\tau\right)\right\|\leqslant\sum_{j=2}^{m(n,\eta)-n}\frac{1}{j!}(\eta(\operatorname{tr}KK^\tau+\delta)+2\delta)^j$$

$$\leqslant(\eta(\operatorname{tr}KK^\tau+\delta)+2\delta)^2e^{\eta(\operatorname{tr}KK^\tau+\delta)+2\delta}. \tag{7.3.12}$$

我们可以选择 η 和 δ 足够小，使得

$$(\eta(\operatorname{tr}KK^\tau+\delta)+2\delta)^2\leqslant\eta\lambda/9, \tag{7.3.13}$$

$$(\eta+3)\delta\leqslant\eta\lambda/6. \tag{7.3.14}$$

于是从式(7.3.12)和(7.3.13)

$$\left\|H_n^{m(n,\eta)-1}\left(-\frac{a}{j}z_jz_j^\tau\right)\right\|\leqslant\frac{1}{9}\lambda\eta e\leqslant\frac{1}{3}\eta\lambda. \tag{7.3.15}$$

现在我们来估计式(7.3.4)右端第二项. 类似式(7.3.9)，

$$\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}z_jz_j^\tau=\sum_{j=n-1}^{m(n,\eta)-2}\frac{a}{j(j+1)}S_j+\frac{a}{m(n,\eta)-1}S_{m(n,\eta)-1}$$

$$-\frac{a}{n-1}S_{n-1}. \tag{7.3.16}$$

故由式(7.3.16)、(7.3.10)和(7.3.14)得

$$\left\|\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}z_jz_j^\tau-\eta KK^\tau\right\|\leqslant\left\|\sum_{j=n-1}^{m(n,\eta)-2}\frac{a}{j(j+1)}S_j\right.$$

$$\left.-\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}KK^\tau\right\|+\frac{2a}{m(n,\eta)}\|KK^\tau\|$$

$$+\left\|\frac{a}{m(n,\eta)-1}S_{m(n,\eta)-1}-aKK^\tau\right\|$$

$$+\left\|\frac{a}{n-1}S_{n-1}-aKK^\tau\right\|\leqslant\left|\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}\left(\frac{S_{j-1}}{j-1}-KK^\tau\right)\right\|$$

$$+\delta+\delta+\delta\leqslant\eta\delta+3\delta\leqslant\eta\lambda/6. \tag{7.3.17}$$

综合式(7.3.4)、(7.3.15)和(7.3.17)，我们有

$$\left\|\prod_{j=n}^{m(n,\eta)-1}\left(I-\frac{a}{j}z_jz_j^\tau\right)\right\|\leqslant\|I-\eta KK^\tau\|$$

$$+\left\|\sum_{j=n}^{m(n,\eta)-1}\frac{a}{j}z_jz_j^\tau-\eta KK^\tau\right\|+\left\|H_n^{m(n,\eta)-1}\left(-\frac{a}{j}z_jz_j^\tau\right)\right\|$$

$$\leqslant1-\eta\lambda+\frac{1}{6}\eta\lambda+\frac{1}{3}\eta\lambda=1-\frac{1}{2}\eta\lambda. \tag{7.3.18}$$

注意到式(7.3.10)， $\frac{a}{j}\|z_j\|^2\leqslant\frac{a}{j}\operatorname{tr}(S_j-S_{j-1})\leqslant2\delta<1$，于是

当$j\geqslant n_0$, $\left\|I-\frac{a}{j}z_jz_j^\tau\right\|<1$, 然后由式(7.3.18), 用类似于上节引理 7.2.3 的方法, 可证明本引理成立. □

引理 7.3.2　在引理 7.3.1 的条件下, 当 $n\geqslant K_n$ 时,

$$\|U_{mn-1}\|\leqslant\left(\frac{n}{m+1}\right)^{\frac{\lambda a}{2}}.$$

证明　由引理 7.3.1

$$\|U_{mn-1}\|\leqslant\exp\left(-\sum_{j=n}^{m}\frac{a\lambda}{2j}\right)$$

$$\leqslant\exp\left[(\log n-\log(m+1))\frac{\lambda a}{2}\right]\leqslant\left(\frac{n}{m+1}\right)^{\frac{\lambda a}{2}}. \qquad □$$

我们继续定理 7.3.1 的证明. 由式(7.3.2)可知

$$\|Q_{m+1}-Q^*\|\leqslant\|U_{mn-1}\|\,\|Q_n-Q^*\|+\left\|\sum_{j=n}^{m}\frac{a}{j}U_{mj}(G_j-z_jz_j^\tau Q^*)\right\|. \qquad (7.3.19)$$

由引理 7.3.1, 上式右端第一项对任意固定的 n, 随 m 趋于无穷而趋于 0. 因而, 为证明定理的充分性, 只需考察第二项.

记

$$T_n=\sum_{j=1}^{n}(G_j-z_jz_j^\tau Q^*). \qquad (7.3.20)$$

由式(7.2.16)中 G_n 的定义及 A7.3.1 和 A7.3.2, 类似上节引理 7.2.1 容易得到

$$\lim_{n}\frac{1}{n}\sum_{j=1}^{n}(G_j-z_jz_j^\tau Q^*)=0,$$

因而对任 $\delta>0$, 存在 n_1, 当 $n\geqslant n_1+1$ 时

$$\frac{a}{n}\|T_n\|<\delta. \qquad (7.3.21)$$

应用分部求和法, 我们可重写式(7.3.19)右边第二项:

$$\left\|\sum_{j=n}^{m}\frac{a}{j}U_{mj}(G_j-z_jz_j^\tau Q^*)\right\|=\left\|\sum_{j=n}^{m}\frac{a}{j}U_{mj}(T_j-T_{j-1})\right\|$$

$$=\left\|\sum_{j=n-1}^{m-1}a\left(\frac{U_{mj}T_j}{j}-\frac{U_{mj+1}T_j}{j+1}\right)+\frac{a}{m}U_{mm}T_m\right.$$

$$-\frac{a}{n-1}U_{mn-1}T_{n-1}\Big\| = \Big\|\sum_{j=n-1}^{m-1} a\left[\left(I-\frac{a}{j+1}z_{j+1}z_{j+1}^{\tau}\right)\Big/ j\right.$$

$$\left.-\frac{1}{j+1}I\right]U_{mj+1}T_j+\frac{a}{m}T_m-\frac{a}{n-1}U_{mn-1}T_{n-1}\Big\|$$

$$\leqslant\Big\|\sum_{j=n-1}^{m-1}\frac{aU_{mj+1}T_j}{j(j+1)}\Big\|+\Big\|\sum_{j=n-1}^{m-1}\frac{a^2 z_{j+1}z_{j+1}^{\tau}U_{mj+1}T_j}{j(j+1)}\Big\|+2\delta$$

$$\leqslant\delta(1+a)\Big(\sum_{j=n-1}^{m-1}\frac{1}{j+1}\|U_{mj+1}\|$$

$$+\sum_{j=n-1}^{m-1}\frac{1}{j+1}\|z_{j+1}\|^2\|U_{mj+1}\|\Big)+2\delta. \qquad (7.3.22)$$

现在，注意到δ的任意性，为完成充分性的证明，只需证明上式括号中的两项一致有界.由引理7.3.1或7.3.2容易证明，括号中的第一项一致有界(在第2、3章中已多次遇到)．用类似于定理7.2.1中的处理方法和式(7.3.17)的估计由引理7.3.2、式(7.3.8)和分部求和知

$$\sum_{j=n-1}^{m-1}\frac{1}{j+1}\|z_{j+1}\|^2\|U_{mj+1}\|\leqslant\sum_{j=n-1}^{m'}\frac{1}{j+1}\left(\frac{j+2}{m+1}\right)^{\frac{1}{2}\lambda a}\|z_{j+1}\|^2$$

$$+\sum_{j=m'+1}^{m-1}\frac{1}{j}\|z_{j+1}\|^2\|U_{mj+1}\|\leqslant 4\left(\frac{1}{m+1}\right)^{\frac{1}{2}\lambda a}$$

$$\cdot\sum_{j=n-1}^{m-1}(j+1)^{\frac{1}{2}\lambda a-1}(\operatorname{tr}S_{j+1}-\operatorname{tr}S_j)+c_1$$

$$\leqslant 4\left(\frac{1}{m}\right)^{\frac{1}{2}\lambda a}\Big(\sum_{j=n-2}^{m-2}[(j+1)^{\frac{1}{2}\lambda a-1}-(j+2)^{\frac{1}{2}\lambda a-1}]\operatorname{tr}S_{j+1}$$

$$+(m-1)^{\frac{1}{2}\lambda a-1}\operatorname{tr}S_{m-1}-n^{\frac{1}{2}\lambda a-1}\operatorname{tr}S_{n-1}\Big)+c_1, \qquad (7.3.23)$$

其中$m'=\min\left\{k:\sum_{j=k+1}^{m-1}\frac{1}{j}\leqslant\eta\right\}$, c_1为正常数.

再利用微分中值公式和$\lim\limits_{n}\frac{1}{n}S_n=KK^{\tau}$,就有

$$\Big|\sum_{j=n-2}^{m-2}[(j+1)^{\frac{1}{2}\lambda a-1}-(j+2)^{\frac{1}{2}\lambda a-1}]\left(\frac{1}{m}\right)^{\frac{1}{2}\lambda a}\operatorname{tr}S_{j+1}\Big|$$

$$\leqslant\sum_{j=n-2}^{m-2}|(j+1)^{\frac{1}{2}\lambda a-1}-(j+2)^{\frac{1}{2}\lambda a-1}|\left(\frac{1}{m}\right)^{\frac{1}{2}\lambda a}\operatorname{tr}S_{j+1}$$

$$\leqslant\left|\frac{1}{2}\lambda a-1\right|\sum_{j=n-2}^{m-2}\left[(j+1)^{\frac{1}{2}\lambda a-2}+(j+2)^{\frac{1}{2}\lambda a-2}\right]\left(\frac{1}{m}\right)^{\frac{1}{2}\lambda a}\operatorname{tr}S_{j+1}$$

$$\leqslant 2\|K\|_*^2\left|\frac{1}{2}\lambda a-1\right|$$

$$\cdot\sum_{j=n-2}^{m-2}\left(\frac{(j+1)^{\frac{1}{2}\lambda a-1}}{m^{\frac{1}{2}\lambda a-1}}+\frac{(j+2)^{\frac{1}{2}\lambda a-1}}{m^{\frac{1}{2}\lambda a-1}}\right)\cdot\frac{1}{m}$$

$$\leqslant 4\|K\|_*^2\left|\frac{1}{2}\lambda a-1\right|\int_0^1 x^{\frac{1}{2}\lambda a-1}dx=\frac{8}{\lambda a}\|K\|_*^2\left|\frac{1}{2}\lambda a-1\right|. \tag{7.3.24}$$

综合式(7.3.23)、(7.3.24)和上述分析，充分性得证.

(ii) 必要性：设 A7.3.1 成立，$Q_n\xrightarrow[n]{}Q^*$. 类似式(7.2.15)，便有

$$Q_{n+1}=Q_n-\frac{a}{n}(z_nz_n^\tau Q_n-G_n), \tag{7.3.25}$$

$$\begin{aligned}nQ_{n+1}&=nQ_n-a(z_nz_n^\tau Q_n-G_n)\\&=Q_n+(n-1)Q_n-a(z_nz_n^\tau Q_n-G_n)\\&=\sum_{j=1}^nQ_j-a\sum_{j=1}^n(z_jz_j^\tau Q_j-G_j),\end{aligned} \tag{7.3.26}$$

故

$$\frac{1}{n}\sum_{j=1}^n(z_jz_j^\tau Q_j-G_j)\xrightarrow[n]{}0. \tag{7.3.27}$$

现往证

$$\frac{1}{n}\left\|\sum_{j=1}^n z_jz_j^\tau(Q_j-Q^*)\right\|_*\xrightarrow[n]{}0. \tag{7.3.28}$$

由于 $Q_n\xrightarrow[n]{}Q^*$，并且 A7.3.1 成立，对任意 $\varepsilon>0$，定存在 n_0，使得当 $n\geqslant n_0$ 时

$$\|Q_n-Q^*\|_*<\frac{1}{4\|K\|_*^2}\varepsilon, \tag{7.3.29}$$

$$\frac{1}{n}\sum_{j=1}^n\|z_j\|_*^2\leqslant 2\|K\|_*^2. \tag{7.3.30}$$

于是

$$\frac{1}{n}\sum_{j=n_0}^{n}\|z_j\|_*^2\|Q_j-Q^*\|_*\leqslant 2\|K\|_*^2\cdot\frac{1}{4\|K\|_*^2}\varepsilon=\frac{1}{2}\varepsilon, \tag{7.3.31}$$

且对这个 n_0，存在 $n_1>n_0$，当 $n\geqslant n_1$ 时，

$$\frac{1}{n}\sum_{j=1}^{n_0-1}\|z_j\|_*^2\|Q_j-Q^*\|_*<\frac{1}{2}\varepsilon. \tag{7.3.32}$$

再从式(7.3.31)和(7.3.32)，对任 $n\geqslant n_1$，

$$\frac{1}{n}\sum_{j=1}^{n}\|z_j\|_*^2\|Q_j-Q^*\|_*\leqslant\varepsilon,$$

故式(7.3.28)成立，再注意式(7.3.27)，即有

$$\frac{1}{n}\sum_{j=1}^{n}(z_jz_j^\tau Q^*-G_j)\xrightarrow[n]{}0. \tag{7.3.33}$$

类似于上节引理 7.2.1 的证明，我们可写

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}(z_jz_j^\tau Q^*-G_j)\\&\quad=\frac{1}{n}\sum_{j=1}^{n}V_2^\tau[Py_jy_j^\tau P(PHH^\tau P)^+(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau)\\&\qquad-P(y_j\psi_j^\tau-y_jy_j^\tau C^{+\tau}\Phi^\tau)+P(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau)\\&\qquad-PHH^\tau P(PHH^\tau P)^+(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau)].\end{aligned} \tag{7.3.34}$$

而由 A7.3.1 和上式，以及式(7.3.33)，可推出

$$\frac{1}{n}V_2^\tau P\sum_{j=1}^{n}(y_j\psi_j^\tau-HL^\tau)\xrightarrow[n]{}0. \tag{7.3.35}$$

由上节知，$Py_j\in L(PH)$ 和 V_2 定义为 $L(PH)$ 的一组基，因此 $Py_j=V_2K_j$，$PH=V_2K$，代入上式得

$$\frac{1}{n}\sum_{j=1}^{n}(K_j\psi_j^\tau-KL^\tau)\xrightarrow[n]{}0,$$

因而

$$\frac{1}{n}P\sum_{j=1}^{n}(y_j\psi_j^\tau-HL^\tau)\xrightarrow[n]{}0,$$

即 A7.3.2 成立，必要性得证. □

定理 7.3.1 给出了当步长因子为 $\left\{\frac{a}{n}\right\}$ 时，算法(7.3.1)a.s. 收

敛的充分必要条件，但这毕竟是一种理想的情况．实际问题中常常并不严格满足这样的充分必要条件，因此，研究在 A7.3.3 和 A7.3.4 成立时，算法的稳健性问题有重要的实际意义．

条件 A7.3.3 和 A7.3.4 中的 ε 刻划了信号与严格的 a.s. 收敛充要条件之间的偏差．在算法的稳健性分析中，我们希望知道为保证算法不发散，这样的偏差容许范围多大；算法与欲求的最小方差滤波阵之间的偏差又如何依赖于 ε．下面的定理将回答上述问题．

定理 7.3.2 如果假设 A7.3.3 和 A7.3.4 成立，而且其中的 $\varepsilon<\min\left(\frac{1}{4},\ \lambda,\ \dfrac{\lambda^2}{\alpha r\left(2\lambda+5\,\mathrm{tr}\,KK^\tau+1\frac{3}{4}\right)^2}\right)$，这里 r 为 HH^τ 的阶数．则算法(7.3.1)有下面的渐近性质：

$$\limsup_n\|x_n-x^*\|\leqslant c_1\varepsilon\quad \text{a.s.} \tag{7.3.36}$$

其中的 c_1 是仅依赖于矩阵 Φ，HH^τ，C 的常数，为记号上的简单，这里未写出具体表达式，但可从证明过程中看出．

证明 为得到(7.3.36)式，一个关键点是分析偏差 ε 应小于什么上界仍有类似于式(7.3.3)的不等式成立．我们仍从式(7.3.2)出发．

首先注意，在 A7.3.3 和 A7.3.4 下，类似式(7.3.34)

$$|\mathrm{tr}\,S_n/n-\mathrm{tr}\,KK^\tau|\leqslant|\mathrm{tr}\,V_2^\tau\varepsilon_nV_2|\leqslant|\mathrm{tr}\,\varepsilon_n|\leqslant r\varepsilon, \tag{7.3.37}$$

$$\begin{aligned}\left\|\frac{1}{n}T_n\right\|&=\left\|\frac{1}{n}\sum_{j=1}^n(G_j-z_jz_j^\tau Q^*)\right\|\\&\leqslant\|V_2\|\|\varepsilon_n\|\|Q^*\|+\|V_2\|\|\eta_n\|+\|V_2\|\|\varepsilon_n\|\|C^{+\tau}\Phi^\tau\|\\&\leqslant C_2\varepsilon,\end{aligned} \tag{7.3.38}$$

其中 r 显然大于等于 S_n 的阶数，而 S_n 和 T_n 定义如式(7.3.8)和(7.3.20)所示，于是类似于式(7.3.9)～(7.3.12)的推理，并用 $\alpha r\varepsilon$ 代替式(7.3.12)中的 δ，我们有

$$\left\| H_n^{m(n,\eta)}\left(-\frac{a}{j} z_j z_j^\tau\right)\right\| \leqslant (\eta \operatorname{tr} KK^\tau + a r\varepsilon(\eta+2))^2 e^{\eta \operatorname{tr} KK^\tau + a r\varepsilon(\eta+2)}. \tag{7.3.39}$$

同样，类似于式(7.3.17)，还有

$$\left\| \sum_{j=n}^{m(n,\eta)-1} \frac{a}{j} z_j z_j^\tau - \eta KK^\tau \right\| \leqslant a r\varepsilon(\eta+3) \tag{7.3.40}$$

进而由式(7.3.4)、(7.3.39)和(7.3.40)，类似于式(7.3.18)的推导，可得

$$\left\| \prod_{j=n}^{m(n,\eta)-1} \left(I - \frac{a}{j} z_j z_j^\tau \right) \right\| \leqslant 1 - \eta\lambda + a r\varepsilon(\eta+3)$$
$$+ (\eta \operatorname{tr} KK^\tau + a r\varepsilon(\eta+2))^2 e^{\eta \operatorname{tr} KK^\tau + a r\varepsilon(\eta+2)}. \tag{7.3.41}$$

由已知条件，$\varepsilon < \min\left(\lambda, \dfrac{1}{4}\right)$和

$$\varepsilon < \frac{\lambda^2}{a r\left(2\lambda + 5 \operatorname{tr} KK^\tau + 1\frac{3}{4}\right)^2},$$

可见

$$2\sqrt{a r\varepsilon}\lambda + 5\sqrt{a r\varepsilon} \operatorname{tr} KK^\tau + 8 a r\varepsilon\sqrt{a r\varepsilon} + \sqrt{a r\varepsilon} < \lambda, \tag{7.3.42}$$

再将式(7.3.42)两端同乘以 $\sqrt{a r\varepsilon}$ 后经整理，可得

$$\frac{\sqrt{a r\varepsilon} - 2 a r\varepsilon}{\operatorname{tr} KK^\tau + a r\varepsilon} > \frac{5 a r\varepsilon}{\lambda - a r\varepsilon}, \quad a r\varepsilon < \min\left(\frac{1}{4}, \lambda\right). \tag{7.3.43}$$

于是我们可选 $\eta = \dfrac{1}{2}\left(\dfrac{\sqrt{a r\varepsilon} - 2 a r\varepsilon}{\operatorname{tr} KK^\tau + a r\varepsilon} - \dfrac{5 a r\varepsilon}{\lambda - a r\varepsilon}\right)$，显然，

$$\frac{5 a r\varepsilon}{\lambda - a r\varepsilon} < \eta < \frac{\sqrt{a r\varepsilon} - 2 a r\varepsilon}{\operatorname{tr} KK^\tau - a r\varepsilon}, \tag{7.3.44}$$

进而

$$(\eta \operatorname{tr} KK^\tau + a\eta r\varepsilon + 2 a r\varepsilon)^2 < a r\varepsilon, \quad \eta > \frac{5 a r\varepsilon}{\lambda - a r\varepsilon}. \tag{7.3.45}$$

因而

$$e^{\eta \operatorname{tr} KK^{\tau}+are(\eta+2)}<e^{\sqrt{are}}<e^{\frac{1}{2}}<2. \qquad (7.3.46)$$

注意式(7.3.44)和 $\lambda<\operatorname{tr} KK^{\tau}$，我们有

$$\frac{5are}{\lambda-are}<\eta<\frac{\sqrt{are}-2are}{\operatorname{tr} KK^{\tau}-are}<\frac{1+5are}{\lambda-are}.$$

因此

$$0<\eta(\lambda-are)-5are<1. \qquad (7.3.47)$$

综合式(7.3.41)、(7.3.45)、(7.3.46)和(7.3.47)，有

$$\begin{aligned}\left\|\prod_{j=n}^{m(n,\eta)-1}\left(I-\frac{a}{j}z_jz_j^{\tau}\right)\right\| &\leqslant 1-\eta\lambda+are(\eta+3)+2are\\ &=1-\eta(\lambda-are)+5are<1.\end{aligned} \qquad (7.3.48)$$

因此，若记 $\alpha=\eta(\lambda-are)-5aer$，则

$$\left\|\prod_{j=n}^{m(n,\eta)-1}\left(I-\frac{a}{j}z_jz_j^{\tau}\right)\right\|\leqslant 1-\alpha<1. \qquad (7.3.49)$$

现在注意到式(7.3.49)和(7.3.38)与式(7.3.18)和(7.3.21)的对应性，以及其中的 $\frac{2\alpha}{\eta}$ 和 $c_2\varepsilon$ 分别对应于前面的 λ 和 δ，完全类似于式(7.3.18)到式(7.3.24)的推理和演算，可以得到如下的不等式和极限

$$\begin{aligned}&\left\|\sum_{j=n}^{m}\frac{a}{j}U_{mj}(G_j-z_jz_j^{\tau}Q^*)\right\|\\ &\leqslant c_2\varepsilon(1+a)\left(c_4+\frac{4\eta}{a\alpha}(1+\|K\|^2)\right)\left|\frac{1}{2}\cdot\frac{2\alpha}{\eta}a-1\right|+2c_2\varepsilon\\ &\leqslant c_1\varepsilon,\end{aligned}$$

$$\|U_{mn-1}(Q_n-Q^*)\|\xrightarrow[m\to\infty]{}0.$$

这也就说明本定理成立. □

在本定理中，我们给出了算法的渐近偏差如何依赖于算法收敛性条件偏差的 ε 的数学描述.但是读者容易看出，这里 ε 的上界和算法渐近偏差上界都依赖于定理推导的实际处理，因此，这些上界是可以改进的.

§7.4 均方收敛与渐近正态性

在前两节中，我们一直是在几乎处处的意义下讨论算法的大样本性质. 在本节，我们将在均方意义下给出算法的收敛性，进而得到算法的渐近正态性. 这不仅具有理论上的意义，而且有重要的实际意义. 因为当算法经过某种正则化以后具有渐近正态性，就可以按实际问题的需要作统计检验，如可以在小样本情况粗略地估计算法在某个偏差范围内的概率. 在这一节我们就将分析适应性波束形成器类算法的渐近正态性. 据作者所知，很多随机递推算法的渐近正态性问题都已得到研究，如随机逼近算法(见前面第3、4章)，最小二乘算法[73, 90]的这类问题已有丰富的结果. 但是，对适应性滤波算法，特别是对适应性波束形成器算法所得到的渐近正态性结果较少；这可能有两方面的原因：第一，在随机逼近算法的渐近正态性问题中，回归函数的线性主部(或梯度阵)应是可逆的，这个要求相应于这里，即要求 $PHH^\tau P$ 可逆，这在阵 $C\neq 0$ 时(见§7.2)，由于 P 阵不满秩而不可能，所以要用我们前面的投影方法把算法变形，使 $PHH^\tau P$ 投影到 $L(PH)$ 子空间上，成为 $V_2^\tau PHH^\tau PV_2$，再在获得渐近正态性结果后利用 Cramer-Wold 的做法[5]把结果变回到原算法. 第二，在适应性波束形成器算法渐近正态性问题研究中，将会出现 $(x_nx_n^\tau-KK^\tau)(Q_n-Q^*)$ 项，这对应于随机逼近渐近正态性问题中的 $\delta(x_n)=o(\|x_n-x^0\|)$ 项，然而前者并不具有 $o(\|Q_n-Q^*\|)$ 的性质，这也给证明它在渐近正态性问题中不起作用带来了困难，后面将会看到，特别当 $\{x_nx_n^\tau\}$ 为无穷相关序列时要证明这一点并不容易. 在本节中，我们将在克服这些困难后对一般的步长因子 $\{a_n\}$ 给出 x_n-x^* 的均方收敛性和 $(x_n-x^*)/\sqrt{a_n}$ 的渐近正态分布.

现在，我们先给出关于均方收敛性所需要的部分假设：

A7.4.1 $a_n>0,\ a_n\xrightarrow[n]{}0,\ a_{n+1}^{-1}-a_n^{-1}\xrightarrow[n]{}\alpha\geqslant 0,\ \sum_{n=1}^{\infty}a_n=\infty.$

A7.4.2 $KK^{\tau}-\frac{1}{2}\alpha I$ 的所有本征值都在右半平面.

A7.4.3 y_n 和 ψ_n 都是 $\mathscr{F}_n$ 可测的，且

$$\sup_n \mathrm{E}(\|y_n\|^4|\mathscr{F}_{n-1})<\infty,\ \sup \mathrm{E}(\|\psi_n\|^4|\mathscr{F}_{n-1})<\infty, \tag{7.4.1}$$

$$\sum_{j=n}^{\infty}a_j\|\mathrm{E}(y_jy_j^{\tau}|\mathscr{F}_{n-1})-HH^{\tau}\|\leqslant ca_n, \tag{7.4.2}$$

$$\sum_{j=n}^{\infty}a_j\|\mathrm{E}(y_j\psi_j^{\tau}|\mathscr{F}_{n-1})-HL^{\tau}\|\leqslant ca_n, \tag{7.4.3}$$

其中常数 c 不依赖于抽样点 ω 和 n，但在不同式于中值可能不同.

注 7.4.1 条件 A7.4.3 在文献[51, 94]中已经用到，并不算限制很强. 首先当 $\{y_ny_n^{\tau}-HH^{\tau}\}$ 和 $\{y_n\psi_n^{\tau}-HL^{\tau}\}$ 为鞅差列时，式(7.4.2)和(7.4.3)显然成立. 其次当

$$\sum_{j=n}^{\infty}\|\mathrm{E}(y_jy_j^{\tau}|\mathscr{F}_{n-1})-HH^{\tau})\|,\ \sum_{j=n}^{\infty}\|\mathrm{E}(y_j\psi_j^{\tau}|\mathscr{F}_{n-1})-HL^{\tau}\|$$

都一致有界时这两式也成立，而这种情况的一些特例在文献[94]中已经给出过. 而且，在后而讨论渐近正态性时，我们还会给出当 $\{y_ny_n^{\tau}\}$ 和 $\{y_n\psi_n^{\tau}\}$ 为某种无穷项滑动平均序列时，也是满足式(7.4.2)和(7.4.3)的例子.

定理 7.4.1 在假设 A7.4.1～A7.4.3 下

$$\mathrm{E}\|x_n-x^*\|^2=O(a_n).$$

证明 和以往一样，我们只需证 $\mathrm{E}\|Q_n-Q^*\|^2=O(a_n)$，为此记 $X_n=(Q_n-Q^*)b$，其中 b 为 R^r 中任一向量，由 b 的任意性，若证得 $\mathrm{E}\|X_n\|^2=O(a_n)$，则可得定理成立.

由式(7.2.22)、(7.4.1)～(7.4.3)，可以得到

$$\begin{aligned}\mathrm{E}(\|X_{n+1}-X_n\|^2|\mathscr{F}_{n-1})\leqslant c\mathrm{E}(&a_n^2\|K\|^4\|X_n\|^2\\&+a_n^2\|V_2^{\tau}P(y_ny_n^{\tau}-HH^{\tau})PV_2X_n\|^2\\&+a_n^2\|V_2^{\tau}P(y_n\psi_n^{\tau}-y_ny_n^{\tau}(C^{+\tau}\Phi^{\tau}\\&+PV_2Q^*))b\|^2/\mathscr{F}_{n-1})\\\leqslant ca_n^2(1+&\|X_n\|^2).\end{aligned} \tag{7.4.4}$$

用上式及恒等式 $\|x\|^2-\|y\|^2=2y^\tau(x-y)+\|x-y\|^2$,

$$
\begin{aligned}
&\mathrm{E}(\|X_{n+1}\|^2-\|X_n\|^2|\mathscr{F}_{n-1})\\
&\leqslant 2\mathrm{E}(X_n^\tau(X_{n+1}-X_n)|\mathscr{F}_{n-1})+ca_n^2(1+\|x_n\|^2)\\
&\leqslant 2X_n^\tau\{-a_nKK^\tau X_n-a_nV_2^\tau PE(y_ny_n^\tau-HH^\tau|\mathscr{F}_{n-1})PV_2X_n\\
&\quad+a_nV_2^\tau P\mathrm{E}(y_n\psi_n^\tau b-y_ny_n^\tau(C^{+\tau}\Phi^\tau+PV_2Q^*)b|\mathscr{F}_{n-1})\}\\
&\quad+ca_n^2(1+\|X_n\|^2). \qquad (7.4.5)
\end{aligned}
$$

容易看出，如果没有上式右端花括号内的第二、三项，问题就容易解决了.为此，我们引进两个扰动辅助函数，使得 $\|X_n\|^2$ 在加上这两个辅助函数后得到的递推不等式中没有类似于式 (7.4.5) 中的那两个"花括号内"项，同时这样的扰动不太大，仍可导出所需要的关于 $\|X_n\|^2$ 的结果. 设

$$
u(n,X)=-2\mathrm{E}\left(\sum_{j=n}^{\infty}(X^\tau a_jV_2^\tau P(y_jy_j^\tau-HH^\tau)PV_2X)|\mathscr{F}_{n-1}\right), \qquad (7.4.6)
$$

$$
\begin{aligned}
\omega(n,X)=2\mathrm{E}\Big(\sum_{j=n}^{\infty}(X^\tau a_jV_2^\tau P(y_j\psi_j^\tau b\\
-y_jy_j^\tau(C^{+\tau}\Phi^\tau+PV_2Q^*)b)|\mathscr{F}_{n-1}\Big), \qquad (7.4.7)
\end{aligned}
$$

由于 X_n 为 $\mathscr{F}_{n-1}$ 可测的，由式(7.4.2)直接可得

$$
\|u(n,X_n)\|\leqslant ca_n\|X_n\|^2. \qquad (7.4.8)
$$

至于对 $\omega(n,X_n)$，我们有

$$
\begin{aligned}
\omega(n,X_n)=&2\sum_{j=n}^{\infty}X_n^\tau a_jV_2^\tau P\\
&\cdot(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau-HH^\tau PV_2Q^*)b\\
&+2X_n^\tau\mathrm{E}\left(\sum_{j=n}^{\infty}(a_jV_2^\tau P(y_j\psi_j^\tau-HL^\tau)b|\mathscr{F}_{n-1}\right)\\
&+2X_n^\tau\mathrm{E}\left(\sum_{j=n}^{\infty}(a_jV_2^\tau P(HH^\tau-y_jy_j^\tau)C^{+\tau}\Phi^\tau b|\mathscr{F}_{n-1}\right)\\
&+2X_n^\tau\mathrm{E}\left(\sum_{j=n}^{\infty}(a_jV_2^\tau P(HH^\tau-y_jy_j^\tau)PV_2Q^*b|\mathscr{F}_{n-1}\right).
\end{aligned} \qquad (7.4.9)
$$

利用式(7.2.10)、(7.2.19)、(7.2.20)可以得到

$$
V_2^\tau PHH^\tau PV_2Q^*=V_2^\tau P(HL^\tau-HH^\tau C^{+\tau}\Phi^\tau), \qquad (7.4.10)
$$

因此，式(7.4.9)右端第一项为0. 再由式(7.4.2)、(7.4.3)和 $a_n\|X_n\|\leqslant a_n(1+\|X_n\|^2)$，可得

$$\|\omega(n, X_n)\|\leqslant ca_n(1+\|X_n\|^2). \tag{7.4.11}$$

现在，我们定义

$$W(n, X)=\|X\|^2+u(n, X)+\omega(n, X). \tag{7.4.12}$$

于是我们有

$$\begin{aligned}
&\mathrm{E}(W(n+1, X_{n+1})|\mathscr{F}_{n-1})-W(n, X_n)\\
&=\mathrm{E}(\|X_{n+1}\|^2-\|X_n\|^2|\mathscr{F}_{n-1})+\mathrm{E}(u(n+1, X_{n+1})|\mathscr{F}_{n-1})\\
&\quad-\mathrm{E}(u(n+1, X_n)|\mathscr{F}_{n-1})+\mathrm{E}(u(n+1, X_n)|\mathscr{F}_{n-1})\\
&\quad-u(n, X_n)+\mathrm{E}(\omega(n+1, X_{n+1})|\mathscr{F}_{n-1})\\
&\quad-\mathrm{E}(\omega(n+1, X_n)|\mathscr{F}_{n-1})+\mathrm{E}(\omega(n+1, X_n)|\mathscr{F}_{n-1})\\
&\quad-\omega(n, X_n),
\end{aligned} \tag{7.4.13}$$

注意上式中

$$\begin{aligned}
&\mathrm{E}(u(n+1, X_n)|\mathscr{F}_{n-1})-u(n, X_n)\\
&=2a_nX_n^{\tau}V_2^{\tau}P\mathrm{E}(y_ny_n^{\tau}-HH^{\tau}|\mathscr{F}_{n-1})PV_2X_n,
\end{aligned} \tag{7.4.14}$$

$$\begin{aligned}
&\mathrm{E}(\omega(n+1, X_n)|\mathscr{F}_{n-1})-\omega(n, X_n)\\
&=-2a_nX_n^{\tau}V_2^{\tau}P\mathrm{E}(y_n\psi_n^{\tau}b-y_ny_n^{\tau}(C^{+\tau}\Phi^{\tau}+PV_2Q^*)b|\mathscr{F}_{n-1}),
\end{aligned} \tag{7.4.15}$$

正是式(7.4.5)右端"花括号"内的第2、3项，但符号相反.

再由类似于式(7.4.4)及式(7.4.8)～(7.4.11)的推理，可以得到

$$\mathrm{E}(\|X_{n+1}-X_n\|/\mathscr{F}_{n-1})\leqslant ca_n(1+\|X_n\|), \tag{7.4.16}$$

$$\begin{aligned}
&\|\mathrm{E}(u(n+1, X_{n+1})-u(n+1, X_n)|\mathscr{F}_{n-1})\|\\
&=\Big\|\mathrm{E}\Big[\sum_{j=n+1}^{\infty}(X_{n+1}^{\tau}a_jV_2^{\tau}P(y_jy_j^{\tau}-HH^{\tau})PV_2X_{n+1}\\
&\quad-X_n^{\tau}a_jV_2^{\tau}P(y_jy_j^{\tau}-HH^{\tau})PV_2X_n)|\mathscr{F}_{n-1}\Big]\\
&=\Big\|\mathrm{E}\Big[(X_{n+1}^{\tau}-X_n^{\tau})\sum_{j=n+1}^{\infty}a_jV_2^{\tau}P(y_jy_j^{\tau}-HH^{\tau})PV_2X_{n+1}|\mathscr{F}_{n-1}\Big]\\
&\quad+\mathrm{E}\Big[X_n^{\tau}\sum_{j=n+1}^{\infty}a_jV_2^{\tau}P(y_jy_j^{\tau}-HH^{\tau})PV_2(X_{n+1}-X_n)|\mathscr{F}_{n-1}\Big]\Big\|
\end{aligned}$$

$$\leqslant ca_{n+1}[\mathrm{E}(\|X_{n+1}-X_n\|(\|X_{n+1}-X_n\|+\|X_n\|)|\mathscr{F}_{n-1})$$
$$+\mathrm{E}(\|X_n\|\|X_{n+1}-x_n\||\mathscr{F}_{n-1})]$$
$$\leqslant ca_{n+1}[ca_n^2(1+\|X_n\|^2)+2ca_n(1+\|X_n\|)\|X_n\|]$$
$$\leqslant ca_{n+1}a_n(1+\|X_n\|^2)\leqslant ca_n^2(1+\|X_n\|^2), \tag{7.4.17}$$

其中最后一个不等号是因为 $1-\frac{a_{n+1}}{a_n}=a_{n+1}(a_{n+1}^{-1}-a_n^{-1})\xrightarrow[n]{}0$.

类似可证

$$\|\mathrm{E}[\omega(n+1, X_{n+1})-\omega(n+1, X_n)|\mathscr{F}_{n-1}]\|$$
$$\leqslant ca_n^2(1+\|X_n\|^2).$$

于是由式(7.4.5)、(7.4.12)～(7.4.17),可以得到

$$\mathrm{E}W(n+1, X_{n+1})-\mathrm{E}W(n, X_n)$$
$$\leqslant -a_n2\lambda\mathrm{E}\|X_n\|^2+Ca_n^2(H\mathrm{E}\|X_n\|^2)$$
$$\leqslant(-a_n2\lambda+Ca_n^2)\mathrm{E}\|X_n\|^2+ca_n^2, \tag{7.4.18}$$

其中 λ 是 KK^τ 的最小本征值,且由 A7.4.2 知 $2\lambda-\alpha>0$.

由式(7.4.8)和(7.4.11)易见

$$-Ka_n+(1-ca_n)\|X_n\|^2\leqslant W(n, X_n)\leqslant ca_n+(1+ca_n)\|x_n\|^2, \tag{7.4.19}$$

于是当 n 足够大以后($ca_n<2\alpha$),由式(7.4.18)和(7.4.19)

$$\mathrm{E}W(n+1, X_{n+1})-\mathrm{E}W(n, X_n)$$
$$\leqslant(-a_n2\lambda+ca_n^2)\mathrm{E}W(n, X_n)+ca_n^2. \tag{7.4.20}$$

再由类似于式(7.2.43)、(7.2.44)的分析,并注意 $2\lambda-\alpha>0$, 可得

$$\mathrm{E}W(n, X_n)=O(a_n), \tag{7.4.21}$$

从不等式(7.4.19),立即有

$$\mathrm{E}|X_n\|^2=O(a_n),$$

故本定理成立. □

为了获得算法的渐近正态性, 我们还需要关于$\{a_n\}$和信号的进一步假设:

A7.4.4 设

$$\frac{\delta_1}{n}\leqslant a_n\leqslant\frac{\delta_2}{\sqrt{n}},\quad 某\ \delta_1>0,\ \delta_2>0, \tag{7.4.22}$$

$$y_n y_n^\tau - HH^\tau = \sum_{i=0}^{\infty} B_i e_{n-i}, \tag{7.4.23}$$

$$P(y_n\psi_n^\tau - y_n y_n^\tau(C^{+\tau}\Phi^\tau - PS^*)) = \sum_{i=0}^{\infty} C_i\omega_{n-i}, \tag{7.4.24}$$

其中 S^* 定义见式(7.2.12)，$(e_n, \mathscr{F}_n)$和$(\omega_n, \mathscr{F}_n)$分别是 $R^{r\times r}$ 和 $R^{r\times m}$ 阶矩阵鞅差序列，$\{B_i\}$和$\{C_i\}$为相容维数的系数阵序列，且

$$\mathrm{E}(e_n|\mathscr{F}_{n-1}) = \mathrm{E}(\omega_n|\mathscr{F}_{n-1}) = 0, \tag{7.4.25}$$

$$\sup_n \mathrm{E}(\|e_n\|^2|\mathscr{F}_{n-1}) < \infty,\ \sup_n \mathrm{E}(\|\omega_n\|^\delta|\mathscr{F}_{n-1}) < \infty, \text{某 } \delta>2. \tag{7.4.26}$$

$$\mathrm{E}(\omega_n\omega_n^\tau|\mathscr{F}_{n-1}) \xrightarrow[n]{} R \quad \text{a.s.}, \tag{7.4.27}$$

$$\sum_{i=0}^{\infty}\|C_i\| < \infty,\ \sum_{j=0}^{\infty} j\|B_j\| < \infty. \tag{7.4.28}$$

注 7.4.2 假设 A7.4.4 进一步给出信号具有如式(7.4.23)和(7.4.24)那样的无穷项滑动平均序列的形式. 这个假设并不与假设 A7.4.3 排斥，例如同时满足式(7.4.2)和式(7.4.26)，则有

$$\sum_{j=n}^{\infty} a_j\|\mathrm{E}(y_j y_j^\tau|\mathscr{F}_{n-1}) - HH^\tau\|$$
$$= \sum_{j=n}^{\infty} a_j\left\|\sum_{i=j+1-n}^{\infty} B_i e_{j-i}\right\| \leqslant ca_n, \tag{7.4.29}$$

而这在$\{e_i\}$有界，$\|B_j\| \leqslant c\dfrac{1}{j^r}$，$r>2$ 时，是成立的，因为

$$\sum_{i=j+1-n}^{\infty}\frac{1}{i^r} \leqslant c\int_{j+1-n}^{\infty}\frac{1}{x^r}\,dx \leqslant c\frac{1}{(j+1-n)^{r-1}},$$

故

$$\sum_{j=n}^{\infty}\left\|\sum_{i=j+1-n}^{\infty} B_i e_{j-i}\right\| \leqslant c. \tag{7.4.30}$$

此外，当$\|B_j\| \leqslant c\dfrac{1}{j^r}$，$r>2$，显然式(7.4.28)成立.

定理 7.4.2 在假设 A7.4.1～A7.4.4 下，有

$$(x_n - x^*)/\sqrt{a_n} \xrightarrow[n\to\infty]{d} N(0, S), \tag{7.4.31}$$

这里 $N(0, S)$表示均值为零阵，协方差阵为 S 的正态分布，其中

$$S=\int_0^\infty e^{Dt}\sum_{i=0}^\infty C_iR\sum_{j=0}^\infty C_j^\tau e^{D^\tau t}\mathrm{d}t, \tag{7.4.32}$$

$$D=-PHH^\tau P+\frac{\alpha}{2}PHH^\tau P(PHH^\tau P)^+. \tag{7.4.33}$$

证明 我们先证 Q_n-Q^* 的渐近正态性. 重写式(7.2.22)为

$$Q_{n+1}-Q^*=(I-a_nKK^\tau)(Q_n-Q^*)$$
$$+a_n(KK^\tau-z_nz_n^\tau)(Q_n-Q^*)+a_n\varepsilon_n, \tag{7.4.34}$$

其中

$$\varepsilon_n=V_2^\tau P(y_n\psi_n^\tau-y_ny_n^\tau(C^{+\tau}\Phi^\tau-PS^*)). \tag{7.4.35}$$

类似式(4.1.12), 我们记

$$\Phi_{n,k}=\begin{cases}\prod_{i=k}^n\left(I-a_i\left(KK^\tau-\frac{1}{2}\alpha I\right)\right), & k\leqslant n,\\ I, & k=n+1.\end{cases} \tag{7.4.36}$$

我们欲证

$$\sum_{k=1}^n\sqrt{a_k}\,\Phi_{n,k+1}(KK^\tau-z_kz_k^\tau)(Q_k-Q^*)\xrightarrow[n]{P}0. \tag{7.4.37}$$

从 A7.4.4 中的式(7.4.23)知, 即需证

$$\sum_{k=1}^n\sqrt{a_k}\,\Phi_{n,k+1}V_2^\tau P\sum_{i=0}^\infty B_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)\xrightarrow[n]{P}0, \tag{7.4.38}$$

$$\sum_{k=1}^n\sqrt{a_k}\,\Phi_{n,k+1}V_2^\tau P\sum_{i=0}^\infty B_ie_kPV_2(Q_k-Q^*)\xrightarrow[n]{P}0. \tag{7.4.39}$$

而由定理 7.4.1, 引理 4.1.2 和 A7.4.4, 并注意 Q_k 是 $\mathscr{F}_{k-1}$ 可测的,

$$\mathrm{E}\left\|\sum_{k=1}^n\sqrt{a_k}\,\Phi_{n,k+1}V_2^\tau P\sum_{i=0}^\infty B_ie_kPV_2(Q_k-Q^*)\right\|^2$$
$$\leqslant c\sum_{k=1}^n a_k\|\Phi_{n,k+1}\|^2\|V_2P\|^2\left(\sum_{i=0}^\infty\|B_i\|\right)^2\mathrm{E}\|Q_k-Q^*\|^2$$
$$\leqslant c\sum_{k=1}^n a_k^2\|\Phi_{n,k+1}\|^2\xrightarrow[n]{}0, \tag{7.4.40}$$

所以要式(7.4.37)成立只需式(7.4.38)成立。

我们重写式(7.4.38)为两项

$$\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}P\sum_{i=0}^{\infty}B_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)$$
$$=\sum_{i=0}^{r-1}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)$$
$$+\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_i(e_{k-i}-e_k)PV_2(Q_k-Q^*).\tag{7.4.41}$$

而上式右端第 2 项又可写成(记 $Q_k=0,\ K<0$)

$$\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)$$
$$=\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_{k-i}PV_2(Q_k-Q_{k-i}+Q_{k-i}-Q^*)$$
$$-\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_kPV_2(Q_k-Q^*).\tag{7.4.42}$$

由式(7.4.25)、(7.4.26)和定理 7.4.1 及 Schwarz 不等式,

$$\mathrm{E}\left\|\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_{k-i}PV_2(Q_{k-i}-Q^*)\right\|$$
$$\leqslant\sum_{i=r}^{\infty}\left(\mathrm{E}\left\|\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_{k-i}PV_2(Q_{k-i}-Q^*)\right\|^2\right)^{\frac{1}{2}}$$
$$\leqslant\sum_{i=r}^{\infty}\|B_i\|\Big(\sum_{k=1}^{n}a_k\|\Phi_{n,k+1}\|^2\|V_2P\|^4$$
$$\cdot\mathrm{E}\|Q_{k-i}-Q^*\|^2\mathrm{E}(\|e_{k-i}\|^2|\mathscr{F}_{k-i-1})\Big)^{\frac{1}{2}}$$
$$\leqslant c\sum_{i=r}^{\infty}\|B_i\|\xrightarrow[r\to\infty]{}0.\tag{7.4.43}$$

$$\mathrm{E}\left\|\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_kPV_2(Q_k-Q^*)\right\|\xrightarrow[r\to\infty]{}0.\tag{7.4.44}$$

然而对式(7.4.42)右端的另一项, 由式(7.4.4)、Schwarz 不等式和式(7.4.28),

$$\mathrm{E}\left\|\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,\Phi_{n,k+1}V_2^{\tau}PB_ie_{k-i}PV_2(Q_k-Q_{k-i})\right\|$$

$$\leqslant\sum_{i=r}^{\infty}\|B_i\|\sum_{k=1}^{n}\sqrt{a_k}\,\|\Phi_{n,k+1}\|\,\|V_2^{\tau}P\|^2\sum_{l=k-i}^{k}\mathrm{E}\|e_{k-i}\|\,\|Q_l-Q_{l-1}\|$$

$$\leqslant c\sum_{i=r}^{\infty}\|B_i\|\sum_{k=1}^{n}\sqrt{a_k}\,\|\Phi_{n,k+1}\|\sum_{l=k-i-1}^{k-1}a_l,\ (\text{记 } a_l=0,\ l<0)$$

$$\leqslant c\sum_{i=r}^{\infty}\|B_i\|\sum_{k=1}^{n}a_k\|\Phi_{n,k+1}\|\sum_{l=k-i-1}^{k-1}\frac{k^{\frac{1}{2}}}{l^{\frac{1}{2}}},\left(\text{记 }\frac{1}{l}=0,\ l<0\right)$$

$$\leqslant c\sum_{i=r}^{\infty}\|B_i\|\sum_{l=1}^{i+1}\frac{(2+i)^{\frac{1}{2}}}{l^{\frac{1}{2}}}\leqslant c\sum_{i=r}^{\infty}\|B_i\|(2+i)\xrightarrow[r\to\infty]{}0. \tag{7.4.45}$$

综合式(7.4.42)～(7.4.45)，我们有

$$\sum_{i=r}^{\infty}\sum_{k=1}^{n}\sqrt{a_k}\,A_{n(k+1)}V_2^{\tau}PB_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)\xrightarrow[r\to\infty]{P}0. \tag{7.4.46}$$

而对固定的正整数 r，注意到

$$\sum_{l=k-i-1}^{k-1}a_l\leqslant Ca_k,\ \forall i\leqslant r,\ \forall k;\ i-1\leqslant k<\infty, \tag{7.4.47}$$

于是

$$\sum_{i=1}^{r-1}\|B_i\|\sum_{k=1}^{n}\sqrt{a_k}\,\|\Phi_{n,k+1}\|\sum_{l=k-i-1}^{k-1}a_l$$

$$\leqslant c\sum_{i=1}^{r-1}\|B_i\|\sum_{k=1}^{n}a_k^{\frac{3}{2}}\|\Phi_{n,k+1}\|\xrightarrow[n]{}0. \tag{7.4.48}$$

利用类似于式(7.4.42)～(7.4.46)的方法及式(7.4.48)容易证明式(7.4.41)右端第一项

$$\mathrm{E}\left\|\sum_{i=0}^{r-1}\sum_{k=1}^{n}\sqrt{a_k}\Phi_{n,k+1}V_2^{\tau}PB_i(e_{k-i}-e_k)PV_2(Q_k-Q^*)\right\|\xrightarrow[n]{}0. \tag{7.4.49}$$

进而由式(7.4.46)和(7.4.49)、(7.4.40)，我们证得式(7.4.37)成立。

类似本书前面§4.1和§4.2关于算法渐近正态性的分析，易

知

$$\frac{1}{\sqrt{a_{n+1}}}(Q_{n+1}-Q^*)=\sum_{k=1}^{n}\sqrt{a_k}\Phi_{n,k+1}\varepsilon_k+o(1), \qquad (7.4.50)$$

其中 $o(1)\xrightarrow[n]{P}0$.

再由式(7.4.24)、(7.4.27)、(7.4.35)和(4.1.31)及定理4.1.1,

$$\frac{1}{\sqrt{a_n}}(Q_n-Q^*)\xrightarrow[n]{d}N(0,\ S'), \qquad (7.4.51)$$

其中

$$S'=\int_0^\infty e^{-Bt}V_2^\tau\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^\tau V_2e^{-Bt}\mathrm{d}t,\quad B=KK^\tau-\frac{\alpha}{2}I. \qquad (7.4.52)$$

仍由 Cramer-Wold 方法[5]并注意 $x_n-x^*=V_2(Q_n-Q^*)$,

$$\frac{1}{\sqrt{a_n}}(x_n-x^*)\xrightarrow[n]{d}N(0,\ S),\ S=V_2S'V_2^\tau. \qquad (7.4.53)$$

为了推出 S 的具体表达式，我们利用 Lyapunov 引理[74]和式(7.4.52)知

$$BS'+S'B=V_2^\tau\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^\tau V_2. \qquad (7.4.54)$$

所以将上式两端同时分别左乘 V_2 和右乘 V_2^τ

$$\begin{aligned}
&V_2\left(V_2^\tau PHH^\tau PV_2-\frac{\alpha}{2}V_2^\tau V_2\right)S'V_2^\tau\\
&\quad+V_2S'\left(V_2^\tau PHH^\tau PV_2-\frac{\alpha}{2}V_2^\tau V_2\right)V_2^\tau\\
&=\left(V_2V_2^\tau PHH^\tau P-\frac{\alpha}{2}V_2V_2^\tau\right)V_2S'V_2^\tau\\
&\quad+V_2S_1'V_2^\tau\left(PHH^\tau PV_2V_2^\tau-\frac{\alpha}{2}V_2V_2'\right)\\
&=V_2V_2^\tau R^*V_2V_2^\tau,
\end{aligned}$$

$$R^*=\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^\tau. \qquad (7.4.55)$$

而由式(7.2.8)和式(7.4.24)知

$$V_2V_2^\tau P=P \quad V_2V_2^\tau R^* V_2V_2^\tau=R^*. \tag{7.4.56}$$

回忆 $PH=V_2K$，于是

$$(PH)^+=K^\tau(KK^\tau)^{-1}(V_2^\tau V_2)^{-1}V_2^\tau=K^\tau(KK^\tau)^{-1}V_2^\tau,$$

即有

$$V_2V_2^\tau=PH(PH)^+. \tag{7.4.57}$$

注意伪逆性质：$A^+=A^\tau(AA^\tau)^+$ 和式(7.4.57)

$$V_2V_2^\tau=PHH^\tau P(PHH^\tau P)^+. \tag{7.4.58}$$

综合式(7.4.55)～(7.4.58)，

$$\begin{aligned}&\left(PHH^\tau P-\frac{\alpha}{2}PHH^\tau P(PHH^\tau P)^+\right)V_2SV_2^\tau\\&\quad+V_2SV_2^\tau\left(PHH^\tau P-\frac{\alpha}{2}PHH^\tau P(PHH^\tau P)^+\right)\\&=R^*.\end{aligned} \tag{7.4.59}$$

于是由式(7.4.53)、(7.4.59)和 Lyapunov 引理知式(7.4.32)和式(7.4.33)成立，定理得证. □

由于 $Ey_ny_n^\tau=HH^\tau$，所以虽然本节的条件和结果都出现了 HH^τ，实际上并不依赖于 H 的选取. 此外，类似于随机逼近算法对一个随机微分方程的逼近，这里也很容易写出类似于注 4.1.1 的同样结果.

§7.5 渐近有效性

和第 4 章类似，在获得算法的渐近正态性以后，自然还希望能得到算法的渐近有效性. 这一节我们将利用与 §4.3 相同的思想和方法，在上节的信号条件下，对步长因子为"慢衰减序列"的适应性滤波算法作平均，我们将证明这样得到的新算法不仅以分布收敛到正态随机阵，而且具有最小的渐近协方差阵.

类似于定理 4.3.1 中随机逼近算法渐近最小方差的导出，易知当 $a_n=\frac{1}{n}$ 时，对适应性滤波算法 $\{Q_n\}$ 也有

$$\min_{\{Q_n\}} \lim_n n\mathrm{E}(Q_n-Q^*)(Q_n-Q^*)^\tau$$

$$=(V_2^\tau PHH^\tau PV_2)^{-1}V_2^\tau \sum_{i=0}^{\infty} C_i R$$

$$\cdot \sum_{i=0}^{\infty} C_i^\tau V_2(V_2^\tau PHH^\tau PV_2)^{-1}. \tag{7.5.1}$$

设 $PH=V_2K$ 为满秩分解，则由伪逆性质知

$$\begin{aligned}(PHH^\tau P)^+ &= (PH)^{+\tau}(PH)^+ \\ &= V_2(KK^\tau)^{-1}KK^\tau(KK^\tau)^{-1}V_2^\tau \\ &= V_2(KK^\tau)^{-1}V_2^\tau = V_2(V_2^\tau V_2KK^\tau V_2^\tau V_2)^{-1}V_2^\tau \\ &= V_2(V_2^\tau PHH^\tau PV_2)^{-1}V_2^\tau. \end{aligned} \tag{7.5.2}$$

再注意 $x_n-x^*=V_2(Q_n-Q^*)$，因此由式(7.5.1)和(7.5.2)，得

$$\min_{\{x_n\}} \lim_n n\mathrm{E}(x_n-x^*)(x_n-x^*)^\tau$$

$$=(PHH^\tau P)^+ \sum_{i=0}^{\infty} C_i R \sum_{i=0}^{\infty} C_i^\tau (PHH^\tau P)^+. \tag{7.5.3}$$

这一段分析说明适应性滤波算法也有类似于随机逼近算法时的渐近最小方差，而且只要算法 x_n 的变形 Q_n 能使式(7.5.1)成立，这样的算法就具有渐近有效性. 下面，我们就来给出这样的算法.

设 $\{x_n\}$ 为式(7.2.1)给出，但其中的增益 $\{a_n\}$ 为 A4.3.1 中式(4.3.5)所定义的慢衰减序列，即

$$a_n>0,\ a_n \xrightarrow[n]{} 0,\ a_n \text{ 非升}, \tag{7.5.4}$$

$$a_{n+1}^{-1}-a_n^{-1}=\frac{a_n-a_{n+1}}{a_na_{n+1}} \xrightarrow[n]{} 0. \tag{7.5.5}$$

我们给出的具有渐近最优性的算法如式(4.3.3)，即

$$\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n} x_i. \tag{7.5.6}$$

易见，这种原算法$\{x_n\}$的平均算法仍然保持

$$\bar{x}_n-x^*=V_2(\bar{Q}_n-Q^*), \tag{7.5.7}$$

这样，我们只需证明法 $\sqrt{n}(\bar{Q}_n-Q^*)$渐近方差为式(7.5.1)等式

右端的阵.

文献[99]在信号为平稳混合序列，步长因子为$\left\{\frac{1}{n^{\alpha}}\right\}$，$\frac{1}{2}<\alpha<1$，$HH^{\tau}>0$，无约束($C=0$，$\Phi=0$)时，运用弱收敛方法证明了算法(7.2.1)在经过平均以后具有渐近有效性．我们在这里将采用§4.3的方法，对更一般形式的适应性滤波算法证明同样的结果.

定理7.5.1 设对由式(7.2.1)、(7.5.4)和(7.5.5)给出的算法$\{x_n\}$，假设A7.4.1～A7.4.4成立，则对算法(7.5.6)，有

$$\sqrt{n}\,(\bar{x}_n-x^*)\xrightarrow[n]{d}N\Big(0,\ (PHH^{\tau}P)^{+}\cdot\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^{\tau}(PHH^{\tau}P)^{+}\Big).$$

证明 正如前面式(7.5.1)～(7.5.7)所分析，我们只需证明对由式(7.2.15)所定义的$\{Q_n\}$成立

$$\sqrt{n}\,(\bar{Q}_n-Q^*)\xrightarrow[n]{d}N\Big(0,\ (V_2^{\tau}PHH^{\tau}PV_2)^{-1}V_2^{\tau}\cdot\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^{\tau}V_2(V_2^{\tau}PHH^{\tau}PV_2)^{-1}V_2\Big).\tag{7.5.8}$$

记

$$u_n=Q_n-Q^*,\ \bar{u}_n=\frac{1}{n}\sum_{i=1}^{n}u_i=\bar{Q}_n-Q^*.\tag{7.5.9}$$

于是由式(7.2.15)知

$$u_{n+1}=(I-a_nKK^{\tau})u_n+a_n(KK^{\tau}-z_nz_n^{\tau})u_n+a_n\varepsilon_n,\tag{7.5.10}$$

其中 $KK^{\tau}=V_2^{\tau}PHH^{\tau}PV_2$，

$$\varepsilon_n=V_2^{\tau}P(y_n\psi_n^{\tau}-y_ny_n^{\tau}(C^{+\tau}\Phi^{\tau}-PS^*)).$$

回忆式(4.3.12)和式(4.3.14)处的记号$\Phi_{n,j}$和$G_{n,j}$，只是用$-KK^{\tau}$代那里的阵H，可以重写上式为

$$u_n=\Phi_{n-1,0}u_0+\sum_{j=0}^{n-1}\Phi_{n-1,j+1}a_j((KK^{\tau}-z_jz_j^{\tau})u_j+\varepsilon_j).\tag{7.5.11}$$

于是由式(7.5.9)和式(4.3.48)知

$$\sqrt{n}\,\bar{u}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{i-1,0}u_0$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=0}^{i-1}\Phi_{i-1,j+1}a_j((KK^\tau-z_jz_j^\tau)u_j+\varepsilon_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{i-1,0}u_0$$

$$+\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\sum_{i=j+1}^{n}\Phi_{i-1,j+1}a_j((KK^\tau-z_jz_j^\tau)u_j+\varepsilon_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{i-1,0}u_0$$

$$+\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}(G_{n,j+1}-(KK^\tau)^{-1})$$

$$\cdot((KK^\tau-z_jz_j^\tau)u_j+\varepsilon_j). \tag{7.5.12}$$

由式(4.3.49)，上式右端第一项趋于0. 而展开第二项可得

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}(G_{n,j+1}-(KK^\tau)^{-1})((KK^\tau-z_jz_j^\tau)u_j+\varepsilon_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}G_{n,j+1}(KK^\tau-z_jz_j^\tau)u_j+\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}G_{n,j+1}\varepsilon_j$$

$$-\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}(KK^\tau)^{-1}\varepsilon_j$$

$$-\frac{1}{\sqrt{n}}(KK^\tau)^{-1}\sum_{j=0}^{n-1}(KK^\tau-z_jz_j^\tau)u_j. \tag{7.5.13}$$

用引理 4.3.1 和类似的方法，我们可证上式右端的第一、二和第四项以概率趋于 0，下而我们仅详细写出第四项以概率趋于 0 的证明.

由式(7.4.23)，可写

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}(KK^\tau-z_jz_j^\tau)u_j$$

$$=\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}V_2^\tau P\Big(\sum_{i=0}^{\infty}B_i(e_{j-i}-e_j+e_j)\Big)PV_2u_j$$

$$=\left(\sum_{i=0}^{r-1}+\sum_{i=r}^{\infty}\right)\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}V_2^{\tau}PB_i(e_{j-i}-e_j+e_j)PV_2u_j.$$

(7.5.14)

类似于式(7.4.43)，易证上式中的

$$\mathrm{E}\left\|\sum_{i=0}^{\infty}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}V_2^{\tau}PB_ie_jPV_2u_j\right\|\xrightarrow[n]{}0. \qquad (7.5.15)$$

由式(7.4.28)，类似于式(7.4.42)～(7.4.44)的推理，并注意 $\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\frac{1}{\sqrt{j}}$ 一致有界，

$$\mathrm{E}\left\|\sum_{i=r}^{\infty}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}V_2^{\tau}PB_ie_{j-i}PV_2(u_j-u_{j-i})\right\|$$

$$\leqslant c\sum_{i=r}^{\infty}\|B_i\|(i+2)\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\frac{1}{\sqrt{j}}\xrightarrow[r\to\infty]{}0. \qquad (7.5.16)$$

于是由式(7.5.15)和式(7.5.16)，可得

$$\mathrm{E}\left\|\sum_{i=r}^{\infty}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}V_2^{\tau}PB_i(e_{j-i}-e_j)PV_2u_j\right\|\xrightarrow[\substack{r\to\infty\\ n\to\infty}]{}0.$$

(7.5.17)

再用类似于式(7.4.47)～(7.4.49)处的推理，可证，对固定的 $r>0$,

$$\sum_{i=0}^{r-1}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}B_ie_{j-i}PV_2u_j\xrightarrow[n]{P}0. \qquad (7.5.18)$$

于是由式(7.5.17)和式(7.5.18)易证式(7.5.13)右端第四项以概率趋于0.

用类似的方法，并注意引理4.3.1的结论2)，可得式(7.5.13)右端第一项以概率趋于0. 同样，可先证式(7.5.13)右端第二项与

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}G_{n,j+1}V_2^{\tau}P\sum_{i=0}^{\infty}C_i\omega_j$$

的差以概率趋于0，然后由式(4.3.16)立即可得

$$\mathrm{E}\left\|\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}G_{n,j+1}V_2^{\tau}P\sum_{i=0}^{\infty}C_i\omega_j\right\|$$

$$\leqslant c\left(\mathrm{E}\frac{1}{n}\sum_{j=0}^{n-1}\|G_{n,j+1}\|^2\|V_2^{\tau}P\|^2\left\|\sum_{i=0}^{\infty}C_i\right\|^2\right)^{\frac{1}{2}}\xrightarrow[n]{}0.$$

经过如上分析,可以写成

$$\sqrt{n}\,\bar{u}_n=\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}(KK^{\tau})^{-1}\varepsilon_j+o(1), \qquad (7.5.19)$$

其中 $o(1)\xrightarrow[n\to\infty]{P}0$.

由于可写

$$\frac{1}{n}\sum_{j=0}^{n-1}\varepsilon_j=\left(1-\frac{1}{n}\right)\left(\frac{1}{n-1}\sum_{j=0}^{n-2}\varepsilon_j\right)+\frac{1}{n}\varepsilon_{n-1}, \qquad (7.5.20)$$

利用定理 4.1.1,立即可得

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\varepsilon_j\xrightarrow[n]{d}N(0,\ S_0), \qquad (7.5.21)$$

其中

$$S_0=\int_0^{\infty}e^{-\frac{1}{2}It}\sum_{i=0}^{\infty}C_iR_i\sum_{i=0}^{\infty}C_i^{\tau}e^{-\frac{1}{2}It}\mathrm{d}t. \qquad (7.5.22)$$

因而

$$(KK^{\tau})^{-1}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\varepsilon_j\xrightarrow[n]{d}N(0,\ (KK^{\tau})^{-1}S_0(KK^{\tau})^{-1}). \qquad (7.5.23)$$

再由 Lyapunov 引理和式(7.5.22),

$$S_0=\frac{1}{2}IS_0+\frac{1}{2}S_0I=\sum_{i=0}^{\infty}C_iR\sum_{i=0}^{\infty}C_i^{\tau}, \qquad (7.5.24)$$

于是从式(7.5.23)和上式,

$$(KK^{\tau})^{-1}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\varepsilon_j\xrightarrow[n]{d}N\Big(0,\ (KK^{\tau})^{-1}\sum_{i=0}^{\infty}C_iR\cdot\sum_{i=0}^{\infty}C_i^{\tau}(KK^{\tau})^{-1}\Big). \qquad (7.5.25)$$

由式(7.5.19)和式(7.5.25),并回忆式(7.5.1)~(7.5.7)的分析,本定理成立。 □

第 8 章

随机并行处理及多传感器数据融合

随着 60 年代各种并行计算机的出现，人们开始研究如何将原有的适合串行计算机的算法加以改造，以充分发挥并行计算机的优势，加快数值计算的速度. 近 30 年来，这种计算机硬件技术和软件技术的发展，相互刺激，相互促进，使得出现了许多专用的、高效的并行处理器，同时并行处理的理论和方法也有了长足的进步(参见参考文献[79]). 但是，直到 80 年代中期，已有的并行处理方法实质上都是确定性的数值计算方法，这些方法当然也可以应用于随机算法的并行处理，因为任何随机算法中都有确定性的数值计算. 然而，随机算法处理的是受到随机扰动的数据，如何压制随机扰动的影响，是改善随机算法性能，加快收敛速度的一个重要问题，这个任务确定性的并行处理方法是不会考虑，也无法完成的. 从 80 年代中期开始，人们开始注意到这一问题，出现了一些从各种具体的随机递推算法的随机特性出发，构造出有效的并行处理方法的研究结果(如参见文献[8, 60, 61, 85]). 此外，在随机信号和随机系统的检测与参数估计问题中，常常采用多传感(观测)器技术来改进处理的精度及稳健性，因此，如何最合理地利用从这些多传感器同时得到的并行数据流，按某种最优准则将它们融合为所需要的一个数据流，是一个十分有意义的问题，国际上近来称这类问题为*多传感器数据融合*(multisensor data fusion)，从某种意义上讲，这也是一类随机并行处理问题. 本章主要介绍随机逼近及适应性滤波两类算法中并行处理和多传感器数据融合的一些研究结果[98, 99, 101, 107, 108, 109, 110]. 对其他类多传感器数据融合感

兴趣的读者还可以参见文献[1, 2].

§8.1 异步并行分布式随机逼近算法

在随机递推算法所处理的实际问题中，常常会遇到实际系统的维数非常大的情况，这时不仅需要计算机有很大的内存，而且按通常串行算法计算耗时也很多，这不仅是因为由维数大带来的计算量大所造成，而且是因为串行计算要求每一分量上都完成同一次迭代后才能进行下一次迭代，这样各分量上的相互等待将造成很多时间延迟. 针对这类问题，文献[85]第一次提出了用若干分离的处理器来处理同一个大系统的各个部分，同时彼此及时地分享处理过程中的信息的思想. 以后，文献[60, 61]又进一步发展了这个思想，提出了一类异步并行分布式的随机逼近算法来解决上述问题. 在他们提出的模型中，人们可用多个处理器来处理同一大系统的各个分量(或子系统，以下不再重复说明)，同时各个处理器之间可以高速地、充分地交流处理信息. 对每一个单个的处理器，在完成一个分量的一次迭代后，立即可以利用这一时刻其它分量上迭代计算的最新信息进行下一步的迭代，不需要等待其它分量也完成同样次数的迭代. 这样，对每一分量，迭代一步所需时间是随机的，到任一确定时刻已迭代的次数也是随机的；而对各个分量来说，算法不同步. 迭代的次数也可能相差甚远. 直观上，这种算法的优点很明显，它不存在各分量之间相互等待造成的时间延迟，在一个固定的时间段里，这种算法在任一分量上都比传统的串行算法所迭代的次数要多. 文献[60, 61]用弱收敛方法从理论上证明了在一定的正则化条件下，这种异步并行随机逼近算法的若干弱收敛结果，数值模拟也显示新方法较原有方法具有更好的收敛性能.

在这一节里，我们将用随机变界截尾办法来处理这类新算法，仍利用 Lyapunov 函数方法来得到它的强收敛性，对进一步的结果仅列出参考文献.

设 $x=(x^1, \cdots, x^l)^\tau \in R^l$，回归函数 $h(\cdot): R^l \to R^l$，为连续函数，其零点集为

$$J=\{x: h(x)=0\}, \tag{8.1.1}$$

后面我们将会看到 $h(\cdot)$ 连续性的要求源于模型复杂性.

对通常的随机逼近算法，我们已假定在一个单位时间内，回归函数 $h(\cdot)$ 在各个分量上都同对完成一次量测并同时完成递推算法的一步迭代. 对这种理想化的情况我们可以用一个处理器同时完成算法在所有分量上所需要的一步观测和迭代. 但是当系统的维数很大时，这种所有分量上的同步性实际上很难做到，于是可设 $\{\tau_n^i\}$ 是一个有界正整数随机序列，τ_n^i 表示在第 i 个分量上完成一步观测和迭代所需的时间. 为了不引起由各分量处理时间不同相互等待造成的时间延误，我们采用 l 个处理器来同时处理一个系统. 因而 τ_n^i 又表示第 i 个处理器在完成 $h(\cdot)$ 的第 i 个分量 $h^i(\cdot)$ 的第 n 次观测和迭代所需的时间. 而第 i 个处理器在完成第 n 次迭代的时刻为 $\{t_n^i\}$，它和 $\{\tau_n^i\}$ 的关系如下：

$$t_0^i=0,\ t_{n+1}^i=t_n^i+\tau_{n+1}^i,\ \forall i\leqslant l. \tag{8.1.2}$$

显然，由 τ_n^i 是正整数值知 $\{t_n^i\}$ 也是一个正整数值的随机变量序列.

设 $\xi^i \in R$，$\xi_{t_n^i}^i$ 是第 i 个处理器在第 n 次迭代中的随机噪声.

设 $x_0=(x_0^1, \cdots, x_0^l)^\tau$ 是回归函数的初始自变量值. 由 $\{t_n^i\}$ 的定义，$x_{t_n^i}^j$ 是自变量的第 j 个分量在第 i 个分量完成第 n 次迭代时的值. 因此，$x_{t_n^i}=(x_{t_n^i}^1, \cdots, x_{t_n^i}^l)^\tau$ 表示自变量在其第 i 个分量完成第 n 次迭代时的值.

若 $n\in[t_j^i, t_{j+1}^i)$，一般 $j\leqslant n$，令

$$x_n^i=x_{t_j^i}^i,\ \xi_n^i=\xi_{t_j^i}^i,\ i\leqslant l. \tag{8.1.3}$$

为了保证各分量的迭代处理不相互等待，我们按分量设计基本算法：

$$x_{t_{n+1}^i}^i=x_{t_n^i}^i+\frac{a}{t_n^i}\left(h^i(x_{t_n^i})+\xi_{t_n^i}^i\right),\ i\leqslant l. \tag{8.1.4}$$

这里的 $a>0$，以后为记号简单，不妨设 $a=1$. 新算法与过去算法

的区别不仅在各分量的迭代时刻是随机的，相互不一定同步，而且在回归函数各分量 $h^i(\cdot)$ 的自变量值在同一处理时刻也很可能是完全不同的.

为了把分量形式的算法写成更紧凑的向量形式的算法，我们需要下列记号：

$$N_i(n)=\max\{k: t_k^i\leq n\},\ i\leq l, \tag{8.1.5}$$

$$\Delta_n^i=n-t_{N_i(n)}^i,\ i\leq l, \tag{8.1.6}$$

$$I_n^i=I_{(\Delta_n^i=0)},\ i\leq l, \tag{8.1.7}$$

其中 $N_i(n)$ 是计数器，它计出到时刻 n 迭代已发生的次数，Δ_n^i 表示到时刻 n 时的最后一次迭代完成后的时间. 若 $\Delta_n^i=0$，则表示到时刻 n 正好完成了一次迭代，这里的 i 均表示所讨论的分量序数.

再定义一个由 l 个 l 维向量组成的集合序列 $\{\tilde{x}_n\}$，其中

$$\tilde{x}_n=\{\tilde{x}_n^1,\ \cdots,\ \tilde{x}_n^l\}, \tag{8.1.8}$$

$$\tilde{x}_n^i=x_{t_{N_i(n)}^i}=(\overset{1}{x}_{t_{N_i(n)}^i},\ \cdots,\ \overset{l}{x}_{t_{N_i(n)}^i})^\tau. \tag{8.1.9}$$

有了这些记号，我们可以重写算法(8.1.4)为

$$x_{n+1}^i=x_n^i+\frac{1}{n}I_{n+1}^i(h^i(\tilde{x}_n^i)+\xi_{n+1}^i),\ i\leq l. \tag{8.1.10}$$

最后定义

$$I_n=\begin{pmatrix} I_n^1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & I_n^l\end{pmatrix}, \tag{8.1.11}$$

$$\xi_n=(\xi_{n+1}^1,\ \cdots,\ \xi_{n+1}^l)^\tau, \tag{8.1.12}$$

$$h(\tilde{x}_n)=(h^1(\tilde{x}_n^1),\ \cdots,\ h^l(\tilde{x}_n^l))^\tau, \tag{8.1.13}$$

其中 $h(\tilde{x}_n)$ 与通常的记号不一样，请读者注意、于是算法(8.1.10)的向量形式为

$$x_{n+1}=x_n+\frac{1}{n}I_{n+1}(h(\tilde{x}_n)+\xi_{n+1}). \tag{8.1.14}$$

虽然式(8.1.14)与通常的 RM 算法形式上很相似，实际上由于对角矩阵 I_{n+1} 因于的增加和 $h(\tilde{x}_n)$ 记号意义上的差别，使新算法实际描述了一个按分量异步并行的随机逼近算法. 对任一时刻 n，若 $I_{n+1}^i=0$，则在第 i 分量上未得到任何新迭代值，$x_{n+1}^i=x_n^i$. 若

$I_{n+1}^i=1$，则按 $x_{n+1}^i=x_n^i+\frac{1}{n}(h^i(\tilde{x}_n^i)+\xi_n^i)$ 得到新的状态值，而且这新值立即传送到其它所有处理器贮存，并在那些处理器进行下一次新的迭代时立即使用.

为获得算法(8.1.14)的强收敛性，我们需要如下假设：

A8.1.1 对任 $i\leqslant l$，设 $\{\tau_n^i\}$ 对 n 一致有界，且满足强大数律，即有

$$\frac{1}{n}\sum_{j=1}^{n}\tau_j^i=\frac{1}{n}t_n^i\xrightarrow[n]{}\mu^i\geqslant 1 \text{ a.s.,} \tag{8.1.15}$$

A8.1.2 对任 $i\leqslant l$，噪声 $\{\xi_{t_n^i}^i\}$ 可分解为

$$\xi_{t_n^i}^i=\phi_{t_n^i}^i+\psi_{t_n^i}^i. \tag{8.1.16}$$

其中

$$\psi_{t_n^i}^i\xrightarrow[n]{}0 \quad \text{a.s.,} \quad \sum_{j=1}^{\infty}\frac{1}{t_j^i}\phi_{t_j^i}^i \quad \text{a.s. 收敛.} \tag{8.1.17}$$

由 $\{\tau_n^i\}$ 的有界性，显然 $\sum_{j=1}^{\infty}\frac{1}{t_j^i}=\infty$，类似可定义

$$\xi_n^i=\phi_n^i+\psi_n^i, \tag{8.1.18}$$

这里的 ϕ_n^i 和 ψ_n^i 可用类似于式(8.1.3)的方式定义.

A8.1.3 存在一个二次连续可微的 Lyapunov 函数 $v(\cdot)$：$R^l\to R^l$，实数 $M>0$ 和一点 $x^*\in\{x:\|x\|<M\}$，使

$$d(v(x),\ v(J))>0,\ \forall x\notin J, \tag{8.1.19}$$

$$v_x^\tau(x)\mu^{-1}h(x)<0,\ \forall x\notin J, \tag{8.1.20}$$

$$v(x^*)<\inf\{v(x):\|x\|=M\}\equiv d, \tag{8.1.21}$$

$$[v(x^*),\ d]\cap v(J)\neq[v(x^*),\ d]. \tag{8.1.22}$$

其中 μ^{-1} 是如下对角矩阵：

$$\mu^{-1}=\begin{pmatrix}\frac{1}{\mu^1} & 0 & \cdots & 0\\ 0 & \frac{1}{\mu^2} & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & \frac{1}{\mu^l}\end{pmatrix},$$

而点 x^* 是以下变界截尾算法中要用到的截尾返回点，

$$d(v(x),\ v(J))\triangleq\inf\{\|v(x)-v(y)\|;\ \forall y\in J\}.$$

设$\{M_n\}$是一单调增至无穷的正实数序列，定义一随机非负整数序列$\{\sigma_n\}$如下：

$$\sigma_0=0,\ \sigma_{n+1}=\sigma_n+I_{\left[\left\|x_n+\frac{1}{n}I_{n+1}(h(\tilde{x}_n)+\xi_n)\right\|>M_{\sigma_n}\right]} \tag{8.1.23}$$

又记

$$J_n=I_{\left[\left\|x_n+\frac{1}{n}I_{n+1}(h(\tilde{x})_n+\xi_n)\right\|<M_{\sigma_n}\right]},$$

$$J_n^c=I_{\left[\left\|x_n+\frac{1}{n}I_{n+1}(h(\tilde{x})_n+\xi_n)\right\|>M_{\sigma_n}\right]},$$

于是定义变界截尾的异步并行随机逼近算法为

$$x_{n+1}=\left(x_n+\frac{1}{n}I_{n+1}(h(\tilde{x}_n)+\xi_n)\right)J_n+x^*J_n^c. \tag{8.1.24}$$

正如前面第3章所说，证明随机逼近算法的有界性是证明算法强收敛性的第一步，而异步并行随机逼近算法又比传统的算法复杂得多，如果不采用变界截尾方法，是很难分析算法的有界性'从而得不到强收敛结果，这再次显示了变界截尾方法是保证算法强收敛性的一个有力工具.

下而，我们先给出一系列引理.

引理8.1.1　在假设A8.1.1下

$$\frac{N_i(n)}{n}\xrightarrow[n]{}\frac{1}{\mu^i},\ \text{a.s.}\quad\forall i\leqslant l$$

证明　由式(8.1.2)和(8.1.5)给出的定义，

$$\frac{t^i_{N_i(n)}}{N_i(n)}\leqslant\frac{n}{N_i(n)}\leqslant\frac{t^i_{N_i(n)+1}}{N_i(n)+1}\cdot\frac{N_i(n)+1}{N_i(n)},\ \forall i\leqslant l.$$

再由A8.1.1，$\frac{N_i(n)+1}{N_i(n)}\xrightarrow[n]{}1$，知引理成立.　□

引理8.1.2　设A8.1.2成立，$\{x_{n_k}\}$由式(8.1.24)所定义的$\{x_n\}$的一个收敛子列，且$\{\tilde{x}_{n_k}\}$有界，则一定存在$\Delta>0$，使得对任η：$0<T<\Delta$，存在k_T，使对$\forall k>k_T$，$\forall m\in[n_k,\ m(n_k,\ T)]$，成立

$$\left\|\sum_{j=n_k}^{m}\frac{1}{j}I_{j+1}(h(\tilde{x}_j)+\xi_j)\right\|\leqslant c\quad\text{a.s.}, \tag{8.1.25}$$

且对$\forall m\in[n_k,\ m(n_k,\ T)+1]$，成立

$$\|x_m - x_{n_k}\| \leqslant c_1 T \quad \text{a.s.}, \tag{8.1.26}$$

由于本引理是§3.1节引理3.1.2的推广，可以照搬那里的证明得到本引理，故此处不再证明.

引理8.1.3 在A8.1.1～A8.1.3下

$$\lim_n \sigma_n \triangleq \sigma < \infty \quad \text{a.s.}.$$

证明 用反证法. 若$\sigma=\infty$，则x_n将无穷多次从x^*出发穿越球$\{x: \|x\|=M\}$. 由A8.1.3知，必存在区间$[\delta_1, \delta_2] \subset [v(x^*), d]$，其中$\delta_1 > v(x^*)$，$\delta_1 \notin v(J)$，而且$v(x_n)$将无穷多次从左边穿越区间$[\delta_1, \delta_2]$. 注意到当$x_n$从$x^*$出发，且$\|x_n\| \leqslant M$时，显然有$\|\tilde{x}_n\| \leqslant M$，因而对这些$x_n$，由A8.1.2和$h(\cdot)$的连续性，

$$x_{n+1} - x_n = \frac{1}{n} I_{n+1}(h(\tilde{x}_n) + \xi_n) \xrightarrow[n]{} 0 \quad \text{a.s.}, \tag{8.1.27}$$

因此我们可从$\{x_n\}$中选出两个子列$\{x_{n_k}\}$和$\{x_{m_k}\}$，满足

$$n_k < m_k, \ \|x_{n_k}\| \leqslant M, \tag{8.1.28}$$

$$v(x_{n_k-1}) < \delta_1, \ \delta_1 \leqslant v(x_j) \leqslant \delta_2, \ n_k \leqslant j \leqslant m_k - 1, \tag{8.1.29}$$

$$v(x_{m_k}) > \delta_2, \tag{8.1.30}$$

且由式(8.1.27)和$v(\cdot)$的连续性，式(8.1.29)中的j是存在的，类似也有

$$\|v(x_{n_k}) - v(x_{n_k-1})\| = \|v_x'(\theta)(x_{n_k} - x_{n_k-1})\| \xrightarrow[k]{} 0 \quad \text{a.s.},$$

于是由式(8.1.29)，

$$v(x_{n_k}) \xrightarrow[k]{} \delta_1. \tag{8.1.31}$$

我们可以从$\{x_{n_k}\}$中再选出一子列，不妨仍记为$\{x_{n_k}\}$，使得

$$x_{n_k} \xrightarrow[k]{} \bar{x} \notin J, \ v(\bar{x}) = \delta_1. \tag{8.1.32}$$

回忆由Taylor展式导出的等式

$$\begin{aligned} v(x_{m(n_k,\eta)+1}) - v(x_{n_k}) &= v_x^{\tau}(\theta)(x_{m(n_k,\eta)+1} - x_{n_k}) \\ &= v_x^{\tau}(\bar{x})(x_{m(n_k,\eta)+1} - x_{n_k}) + (\theta - \bar{x})^{\tau} v_{xx}(\bar{\theta})(x_{m(n_k,\eta)+1} - x_{n_k}). \end{aligned} \tag{8.1.33}$$

由于$v_{xx}(\cdot)$连续，$\|v_{xx}(\bar{\theta})\|$有界，只要$k \geqslant k_T$，引理2和式(8.1.32)又导出当$k \geqslant k_T$，$\|\theta - \bar{x}\| \leqslant 2c_1 T$，因此，

$$
\begin{aligned}
v(x_{m(n_k,T)+1})-v(x_{n_k}) &= v_x^\tau(\bar{x})\,(x_{m(n_k,T)+1}-x_{n_k})+O(T^2)\\
&= v_x^\tau(\bar{x})\sum_{j=n_k}^{m(n_k,T)}\frac{1}{j}\,I_{j+1}(h(\bar{x}_j)+\psi_j+\phi_j)+O(T^2).
\end{aligned}
\tag{8.1.34}
$$

由 A8.1.2 知

$$
v_x^\tau(\bar{x})\sum_{j=n_k}^{m(n_k,T)}\frac{1}{j}\,I_{j+1}(\psi_j+\phi_j)\xrightarrow[k\to\infty]{}0\quad \text{a.s.}, \tag{8.1.35}
$$

再注意到 $h(\tilde{x}_j)$ 的定义，对式(8.1.34)右边第一项，

$$
\begin{aligned}
v_x^\tau(\bar{x})\sum_{j=n_k}^{m(n_k,T)}\frac{1}{j}\,I_{j+1}h(\tilde{x}_j) &= \sum_{i=1}^{l} v_{x^i}(\bar{x})\sum_{j=n_k}^{m(n_k,T)}\frac{1}{j}\,I^i_{j+1}h^i(\tilde{x}^i_j)\\
&= \sum_{i=1}^{l} v_{x^i}(\bar{x})\sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{1}{t^i_j}h^i(x_{t^i_j})\\
&= \sum_{i=1}^{l} v_{x^i}(\bar{x})h^i(\bar{x})\sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{1}{t^i_j}+\sum_{i=1}^{l} v_{x^i}(\bar{x})\\
&\quad\cdot\sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{1}{t^i_j}(h^i(x_{t^i_j})-h^i(\bar{x})).
\end{aligned}
\tag{8.1.36}
$$

由 $h(\cdot)$ 的连续性，和引理 8.1.2 及式 (8.1.32)，对任 $m\in[n_k,\ m(n_k,\ T)]$，我们有

$$
\|x_m-\bar{x}\|\leqslant\|x_m-x_{n_k}\|+\|x_{n_k}-\bar{x}\|\leqslant c_2T, \tag{8.1.37}
$$

$$
\max_{t^i_j\in[N_i(n_k),N_i(m(n_k,T))]}\|h^i(x_{t^i_j})-h^i(\bar{x})\|\xrightarrow[\eta\to 0]{}0, \tag{8.1.38}
$$

因此式(8.1.36)右端第二项为 $o(T)$ 项 $\left(\text{即 } \dfrac{o(T)}{T}\xrightarrow[T\to0]{}0\right)$.

注意到由引理 1 和 $N_i(n)$ 的定义，

$$
\begin{aligned}
\sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{1}{t^i_j} &= \sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{j}{t^i_j}\cdot\frac{1}{j}\\
&= \left(\frac{1}{\mu^i}+o(1)\right)\sum_{j=N_i(n_k)}^{N_i(m(n_k,T))}\frac{1}{j}\\
&= \left(\frac{1}{\mu^i}+o(1)\right)(\log N_i(m(n_k,\ T))-\log N_i(n_k)+o(1))\\
&= \left(\frac{1}{\mu^i}+o(1)\right)\left(\log\frac{m(n_k,\ T)}{\mu^i}-\log\frac{n_k}{\mu^i}+o(1)\right)\\
&= \frac{1}{\mu^i}T+o(1),
\end{aligned}
\tag{8.1.39}
$$

其中 $o(1) \xrightarrow[k\to\infty]{} 0$. 于是对式(8.1.36)右端第一项有

$$\sum_{i=1}^{l} v_{x^i}(\bar{x}) h^i(\bar{x}) \sum_{j=N_i(n_k)}^{N_i(m(n_k,T))} \frac{1}{t_j^i} = T v_x^{\tau}(\bar{x}) \mu^{-1} h(\bar{x}) + o(1). \tag{8.1.40}$$

综合式(8.1.34)～(8.1.40)及式(8.1.32), 由式(8.1.20)知

$$\limsup_k v(x_{m(n_k,T)+1}) \leqslant \delta_1 - c_3 T. \tag{8.1.41}$$

另一方面, 由式(8.1.37)和(8.1.32), $v(\cdot)$一致连续,

$$\max_{n_k \leqslant m \leqslant m(n_k,T)} |v(x_m) - v(x_{n_k})| \xrightarrow[T\to 0]{} 0, \text{一致于 } k \geqslant k_T,$$

这样当 T 适当小, 对任 $m \in [n_k, m(n_k, T)+1)$, $v(x_m) < \delta_2$, 回忆式(8.1.28)～(8.1.30)中对子列$\{x_{n_k}\}$和$\{x_{m_k}\}$的选法, 应有

$$\delta_1 \leqslant v(x_{m(n_k,T)+1}) \leqslant \delta_2,$$

而这与式(8.1.41)矛盾, 故引理得证. □

有了以上三个引理, 再按照§2.4中给出的 Lyapunov 函数方法, 容易得到以下定理:

定理 8.1.1 设 A8.1.1～A8.1.3 成立, 则对任意初值 x_0, 变界截尾异步并行随机逼近算法(8.1.24)是强一致的, 即

$$\lim_n d(x_n, J) = 0 \quad \text{a.s.}.$$

由于定理证明的完成只是照搬§2.4中的推理, 我们不再写出, 读者可以参阅文献[98]. 另外, 关于非加性噪声时算法的强一致性, 关于算法的强收敛速度及渐近正态性, 由于分析的方法与已有的通常的算法相应结果类似, 我们在这里就不再占篇幅, 有兴趣的读者可参阅文献[98, 108].

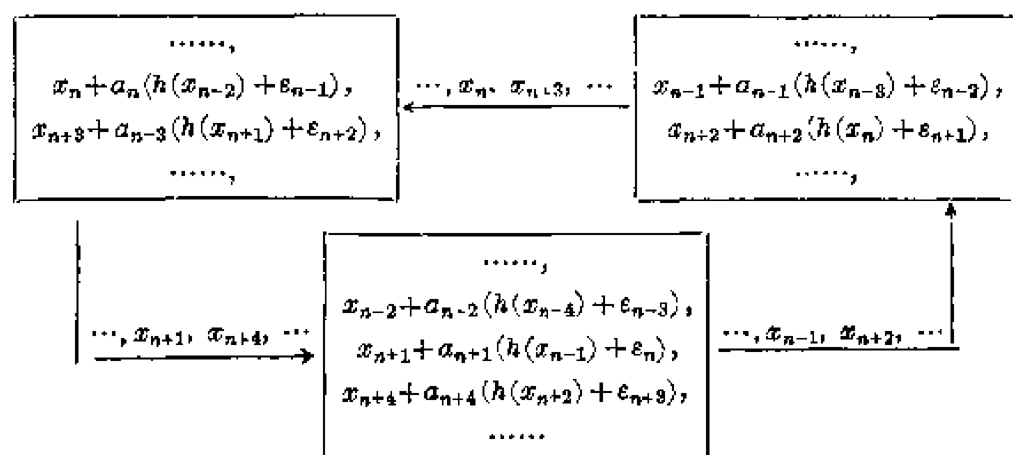
§8.2 多处理器序贯异步并行时滞随机逼近算法

由随机逼近算法明确的实际背景决定了算法的每一步迭代都伴随着对实际系统的状态或参数的调控, 以及对此状态或参数下系统的观测, 这在系统维数较大, 或者系统为遥控目标时费时是较

多的. 这也就是说, 随机逼近算法的一步迭代主要耗时在得到观测量 $y_{n+1}=h(x_n)+\varepsilon_{n+1}$, 而余下的加法和乘法都不与实际系统发生关系, 可在处理器内部很快完成. 因此, 为了缩短随机逼近算法的一步迭代时间, 我们考虑一种多机序贯异步并行的处理方法.

我们通过一个实际例子来说明这种方法的基本思想. 设 x 是一个飞行器的可调控参数, $h(x)$ 是这个飞行器的可观测飞行状态. 人们在地面上进行观测, 发出调控信号调节参数 x, 并使得飞行器的飞行状态去逼近所需要的目标值 $h(\theta)$, 也即使 x 去逼近 θ. 于是, 我们可以采用随机逼近算法递推地完成这一任务. 假设对一个处理器, 从向飞行体发出改变参数 x 的信号, 经过飞行器接收信号并改变参数, 到地面完成对飞行器在新参数下飞行状态 $h(x)$ 的观测共需 r 个时间单位, 设完成随机逼近算法一步迭代还需要的乘法和加法需 1 个时间单位, 因而完成一步迭代总共需 $r+1$ 个时间单位. 这样, 在 $r+1$ 个时间单位里人们只发出 1 次信号, 进行 1 次观测, 飞行器只改变 1 次参数和飞行状态. 现在我们利用 $r+1$ 个同样有上述功能的处理器序贯地进行工作, 虽然对每一单个的处理器同样需要 $r+1$ 个时间单位才能完成一步迭代 $x_{n+1}=x_n+a_n(h(x_{n-r})+\varepsilon_{n-r})$ (注意此时的足标 n 已不表示这个处理器的第 n 次迭代, 而表示第 n 个时间单位时刻), 但若把这 $r+1$ 个处理器构成一个环状相连的并行处理器, 使得每一处理器一步迭代所得到的 x_{n+1} 立即传递给下一个处理器, 使得每一处理器在费时 r 个单位完成观测时正好得到新的参数值. 这样, 在每一个时间单位时刻, 在这一并行处理器中都有处理器在发出信号, 进行观测, 作简单的加法和乘法运算并将结果送给下一处理器, 而飞行器都在改变参数和飞行状态. 这也就相当于对这个由 $r+1$ 个单个处理器构成的并行处理器来说, 一步迭代的时间为 1 个时间单位. 这就是所谓的多机序贯异步并行处理器的工作原理.

为使读者更易理解, 当 $r=2$, 即一步迭代时间为 3 的特例. 用一个简图来说明:



我们还可以考虑更复杂的情况，例如 r 为一个有界随机正整数变量，但这时也可将这种情况简化为 r 取上界的常值情况，这里不再深入下去.

在构造好针对这种情况的并行处理器后，余下的问题是带有时滞的算法

$$x_{n+1}=x_n+a_n(h(x_{n-r})+\varepsilon_{n-r+1}) \tag{8.2.1}$$

的收敛性和渐近性质分析，只有当这个算法仍具有通常算法那样良好的收敛性质，才能说明这种多机序贯并行处理直观上的优越性具有理论上的依据. 为此，我们仍需对算法(8.2.1)作如下的变界截尾改造:

设$\{M_n\}$是一单调趋于无穷的正实数序列，x^* 为 R^l 中某一点，我们递推地定义一个正整数随机序列$\{\sigma_n\}$和算法$\{x_n\}$如下:

$$\sigma_n=\sum_{i=r}^{n-1} I_{[\|x_i+a_i(h(x_{i-r})+\varepsilon_{i-r+1})\|>M_{\sigma_i}]},$$
$$\sigma_0=\sigma_1=\cdots=\sigma_r=0, \tag{8.2.2}$$

$$x_{n+1}=\begin{cases} x_n+a_n(h(x_{n-r})+\varepsilon_{n-r+1}), \\ \quad 当\ \|x_n+a_n(h(x_{n-r})+\varepsilon_{n-r+1})\|\leqslant M_{\sigma_n}, \\ x_{n+2}=x_{n+3}=\cdots=x_{n+r+1}=x^*, \quad 反之, \end{cases} \tag{8.2.3}$$

经过这样构造的算法(8.2.2)和(8.2.3)，可以完全照搬§3.1中的收敛性分析方法，得到算法的有界性和强一致性，于是我们有下面的定理.

定理8.2.1 在与定理3.1.1相同的假设下，对任意初值 x_0,

$x_1, \cdots, x_r$，算法(8.2.2)和(8.2.3)所定义的 x_n 有

$$\lim_{n\to\infty} d(x_n, J)=0 \quad \text{a.s.}.$$

由于本定理的证明与定理 3.1.1 对照，没有任何实质性的差别，我们不再给出具体的证明，关心证明细节的读者可以参阅文献[110].

如果设回归函数 $h(\cdot)$ 在零点 x^0 处连续，并改写式(8.2.1)为

$$x_{n+1}=x_n+a_n(h(x_n)+h(x_{n-r})-h(x_n)+\varepsilon_{n-r+1}), \tag{8.2.4}$$

并把其中的 $h(x_n)-h(x_{n-r})$ 也看成噪声项来分析，则把具有时滞的算法化成了非时滞的算法，这时只要已得到 $x_n \xrightarrow[n]{} x^0$ a.s.，我们可以用类似 § 4.1～4.2 的方法来分析变界截尾时滞算法(8.2.2)和(8.2.3)的强收敛速度和渐近正态性，这时我们也将有与定理 4.2.1 类似的渐近正态性定理[110]给出

$$\frac{x_n-x^0}{\sqrt{a_n}} \xrightarrow[n\to\infty]{d} N(0, S), \lim_n \frac{1}{a_n}\mathrm{E}\|x_n-x^0\|^2=\operatorname{tr} S, \tag{8.2.5}$$

其中 S 为某对称半正定阵．当取 $a_n=\frac{1}{n}$，上式成

$$\sqrt{n}\,(x_n-x^0) \xrightarrow[n\to\infty]{d} N(0, S), \quad \lim_n n\mathrm{E}\|x_n-x^0\|^2=\operatorname{tr} S. \tag{8.2.6}$$

需要特别指出的是这里的 n 既是迭代次数，又是时间单位时刻，然而对单处理器通常算法 $\{\bar{x}_m\}$ 的情况，如果仍用 n 表示时间单位时刻，由于在时刻 n，实际只迭代了 $\left[\frac{n}{r+1}\right]$ 次，那么由渐近正态性定理，应有

$$\sqrt{\left[\frac{n}{r+1}\right]}\left(\bar{x}_{\left[\frac{n}{r+1}\right]}-x^0\right) \xrightarrow[n\to\infty]{d} N(0, S), \tag{8.2.7}$$

也即有

$$\lim_{n\to\infty}\left[\frac{n}{r+1}\right]\mathrm{E}\|\bar{x}_{\left[\frac{n}{r+1}\right]}-x^0\|^2=\operatorname{tr} S, \tag{8.2.8}$$

因此，从上述分析可以得到，要达到同样的逼近精度，单处理器的传统算法所需的时间是多处理器时滞算法的 $r+1$ 倍. 这个效益正是多处理器换来的，类似的思想还可以用于其他递推算法. 文献[68]介绍了在信号处理的实际问题中，也会遇到需采用时滞适应性滤波算法的若干情况，因此，我们可类似地采用多机序贯异步并行处理的技术，并应用如下形式的时滞适应性滤波算法

$$x_{n+1}=C^{+\tau}\Phi^{\tau}+P[x_n+a_{n-r}(y_{n-r}\psi_{n-r}^{\tau}-y_{n-r}y_{n-r}^{\tau}x_{n-r})],$$

$$x_0=x_1=\cdots=x_{r-1}=C^{+\tau}\Phi^{\tau}.$$

为了得到这类算法的强收敛性，我们仍然可以采用本节所用的办法，先将上述算法改造成变界截尾形式，证明修改后的算法一致有界，然后再得到算法强一致性. 我们不再详细讨论，有兴趣的读者可参阅文献[109].

§8.3 有关凸组合优化的矩阵函数极值定理

在后面两节，我们将考虑多传感(观测)器时随机逼近算法与适应性滤波算法的数据融合问题，其中都会遇到一类有凸线性约束的非线性矩阵函数的求极值问题. 因此，在本节我们先给出一个预备性的定理.

设一组矩阵 $A_i\in R^{l\times l}$, $i=1, 2, \cdots, r$ 为对称正定阵，即

$$A_i=A_i^{\tau},\ A_i>0,\quad i=1,2,\cdots,r. \tag{8.3.1}$$

设 $F(\cdot)$是定义于 $R^{l\times l}$，取值于 R 上的关于 $l\times l$ 阶阵可微的非线性函数，我们希望在约束

$$\sum_{i=1}^{r}X_i=I_l\quad (I_l\text{ 为 } l\times l \text{ 阶单位阵}) \tag{8.3.2}$$

下，求出非线性矩阵函数 $F\left(\sum_{i=1}^{r}X_iA_iX_i^{\tau}\right)$的某些(不一定全部)极值. 若采用 Lagrange 乘子法解上述问题，就化为下面非线性矩阵函数方程组的求解问题:

$$\partial\left(F\left(\sum_{j=1}^{r} X_jA_jX_j^{\tau}\right)+\operatorname{tr}\lambda\left(\sum_{j=1}^{r} X_j-I_l\right)\right)\Big/\partial X_i=0,\ i=1,\ 2,\cdots,\ r, \tag{8.3.3}$$

其中 λ 也是待求的 $l\times l$ 阶阵.

根据公式(1.4.31).

$$\frac{\partial\operatorname{tr}\lambda\left(\sum_{j=1}^{r} X_j-I\right)}{\partial X_i}=\lambda^{\tau},\quad i=1,\ 2,\ \cdots,\ r, \tag{8.3.4}$$

我们可重写式(8.3.3)为

$$\frac{\partial F\left(\sum_{j=1}^{r} X_jA_jX_j^{\tau}\right)}{\partial X_i}+\lambda^{\tau}=0,\quad i=1,\ 2,\ \cdots,\ r. \tag{8.3.5}$$

现在我们给出本节的主要结果:

定理8.3.1　在本节的假设下，方程组(8.3.3)总有一组(不一定全部)不依赖于函数 $F(\cdot)$形式的解

$$X_i^*=\left(\sum_{j=1}^{r} A_j^{-1}\right)^{-1}A_i^{-1},\quad i=1,\ 2,\ \cdots,\ r. \tag{8.3.6}$$

证明　我们首先利用§1.4中的矩阵微分公式给出方程(8.3.5)左端的第一项. 由公式(1.4.25):

$$\begin{aligned}
&\frac{\partial F\left(\sum_{j=1}^{r} X_jA_jX_j^{\tau}\right)}{\partial X_i}\\
&\quad=\frac{\partial\operatorname{vec}\left(\sum_{j=1}^{r} X_jA_jX_j^{\tau}\right)^{\tau}}{\partial X_i}\left(I_l\otimes\left.\frac{\partial F(Z)}{\partial\operatorname{vec}Z}\right|_{Z=\sum_{j=1}^{r}X_jA_jX_j^{\tau}}\right)\\
&\quad=\frac{\partial\operatorname{vec}(X_iA_iX_i^{\tau})^{\tau}}{\partial X_i}\left(I_l\otimes\left.\frac{\partial F(Z)}{\partial\operatorname{vec}Z}\right|_{Z=\sum_{j=1}^{r}X_jA_jX_j^{\tau}}\right).
\end{aligned} \tag{8.3.7}$$

再由式(1.4.22)、(1.4.40)和(1.4.42)

$$\begin{aligned}
\frac{\partial\operatorname{vec}(X_iA_iX_i^{\tau})^{\tau}}{\partial X_i}&=\frac{\partial(\operatorname{vec}A_i)^{\tau}(X_i^{\tau}\otimes X_i^{\tau})}{\partial X_i}\\
&=(I_l\otimes(\operatorname{vec}A_i)^{\tau})\ \frac{\partial(X_i\otimes X_i)}{\partial X_i}\\
&=(I_l\otimes(\operatorname{vec}A_i)^{\tau})(U\otimes X_i^{\tau}+(I_l\otimes U)(U\otimes X_i^{\tau})(I_l\otimes U),
\end{aligned} \tag{8.3.8}$$

这里的 U 是 $l\times l$ 阶阵的置换矩阵，于是由式(8.3.3)～(8.3.8)，方程(8.3.3)化为

$$(I_l\otimes(\mathrm{vec}\ A_i)^\tau)(U\otimes X_i^\tau+(I_l\otimes U)(U\otimes X_i)(I_l\otimes U)$$
$$\cdot\left(I_l\otimes\frac{\partial F(Z)}{\partial\ \mathrm{vec}\ Z}\bigg|_{Z=\sum_{j=1}^{r}X_jA_jX_j^\tau}\right)+\lambda^\tau=0. \tag{8.3.9}$$

现在，我们仅需要证明定理给出的一组解(8.3.6)满足下列方程

$$\sum_{i=1}^{r}X_i^*=I, \tag{8.3.10}$$

$$\lambda=-(I_l\otimes(\mathrm{vec}\ A_i)^\tau)(U\otimes X_i^{*\tau}+(I_l\otimes U)(U\otimes X_i^*)$$
$$\cdot(I_l\otimes U)\left(I_l\otimes\frac{\partial F(Z)}{\partial\ \mathrm{vec}\ Z}\bigg|_{Z=\sum_{j=1}^{r}X_j^*A_jX_j^{*\tau}}\right),$$

对 $\forall i=1,\cdots,r$ 成立. (8.3.11)

由式(8.3.1)知，显然式(8.3.10)成立. 余下只需证明对任意的 i，由式(8.3.11)给出的阵 λ 都是与 i 无关的同一常阵，也即式(8.3.6)给出的解组为方程(8.3.3)的解.

由于

$$Z=\sum_{j=1}^{r}X_j^*A_jX_j^{*\tau}=\left(\sum_{j=1}^{r}A_j^{-1}\right)^{-1}, \tag{8.3.12}$$

所以式(8.3.11)右端第三个因子

$$I_l\otimes\frac{\partial F(Z)}{\partial\ \mathrm{vec}\ Z}\bigg|_{Z=\sum_{j=1}^{r}X_j^*A_jX_j^{*\tau}}$$

是一个常阵.

注意到由置换阵的公式(1.4.23)，这里的

$$U=\begin{bmatrix} E_{11} & E_{21} & \cdots & E_{l1}\\ E_{12} & E_{22} & \cdots & E_{l2}\\ & \cdots & & \\ E_{1l} & E_{2l} & \cdots & E_{ll}\end{bmatrix} \tag{8.3.13}$$

其中的 E_{ij} 是 $l\times l$ 阶的初等矩阵. 于是将式(8.3.11)右端的前两项因子的积展开，利用上式，其中第一项为

$$(I_l\otimes(\mathrm{vec}A_i)^\tau)\left(U\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\right)$$

$$=\begin{pmatrix}(\mathrm{vec}\,A_i)^\tau & & & \\ & (\mathrm{vec}\,A_i)^\tau & & 0 \\ 0 & & \ddots & \\ & & & (\mathrm{vec}\,A_i)^\tau\end{pmatrix}$$

$$\cdot\begin{pmatrix}E_{11}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\cdots E_{l1}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\\ E_{12}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\cdots E_{l2}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\\ E_{1l}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\cdots E_{ll}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\end{pmatrix}. \tag{8.3.14}$$

由 A_i 的对称性以及公式(1.4.19)、(1.4.22)和(1.4.17)，上式右端用分块矩阵乘法得到积矩阵中第(k, j)块子矩阵为

$$(\mathrm{Vec}\,A_i)^\tau\left(E_{jk}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\right)$$

$$=(\mathrm{Vec}\,I_l)^\tau(I_l\otimes A_i)\left(E_{jk}\otimes A_i^{-1}\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\right)$$

$$=(\mathrm{Vec}I_l)^\tau\left(E_{jk}\otimes\left(\sum_{j=1}^r A_j^{-1}\right)^{-1}\right),$$

$$i=1,\ 2,\ \cdots,\ l;\ j=1,\ 2,\ \cdots,\ l. \tag{8.3.15}$$

显然，它们均为不依赖于 i 的常阵.

从式(8.3.13)和 U 的定义(1.4.9)，这里的 U 满足

$$U^\tau=U,\ U\ \mathrm{vec}\ A_i=\mathrm{vec}\ A_i^\tau=\mathrm{vec}\ A_i,\ i=1,\ 2,\ \cdots,\ r \tag{8.3.16}$$

因此，对式(8.3.11)右端前两个因子积的第二项，用公式(1.4.17)：

$$(I_l\otimes(\mathrm{vec}\ A_i)^\tau)(I_l\otimes U)(U\otimes X_i^*)(I_l\otimes U)$$

$$=(I_l\otimes(\mathrm{vec}\ A_i)^\tau U^\tau)(U\otimes X_i^*)(I_l\otimes U)$$

$$=(I_l\otimes(\mathrm{vec}\ A_i)^\tau)(U\otimes X_i^*)(I_l\otimes U). \tag{8.3.17}$$

可知上式右端前两项因子的积同式(8.3.14)是不依赖于 i 的常阵，故这第二项也是常阵.

综合式(8.3.12)～(8.3.17)的推理，式(8.3.11)给出的 λ 确实是不依赖于 i 的同一常阵，故定理成立. □

§8.4 多观测器的最优凸组合随机逼近算法

在随机逼近的一些实际问题中，例如前面所举的在地面上观测和调控空中飞行器至最佳参数的例子，人们实际上是在地面上的不同地点用多个观测器观测这个飞行器，从而同时得到多路观测数据，又如 §2.1 中所举的试验一种新药的最佳剂量的例于，人们经常是在多个地区的不同医院里同时试验. 因此，在随机逼近算法的实际应用中，通过多个观测器得到多路数据的情况是大量的. 于是很自然地产生了一个问题：怎样才能合理地、充分地利用所得到的全部信息，构造出在某种意义下最优的算法，而不是凭经验仅利用其中某一路数据或对多路信息作简单的算术平均？

在这一节里，我们将根据不同的实际背景，提出两类不同的并行观测下的凸组合随机逼近算法，从理论上严格证明算法的最优凸组合系数在最小方差意义下是什么，最后还将给出这些最优凸组合系数算法的实现途径.

算法 I

在新药最佳剂量确定的问题中，各个不同的医院完全独立地各自进行着试验，相互之间并不交流信息，也就是说各自独立地按随机逼近算法选择观测点，用观测值完成一步迭代，如此不断进行，得到针对同一问题的一组随机逼近算法 $\{x_n^i\}$，$i\leqslant r$，然后我们用凸组合方式产生一个新的算法序列 $\{z_n\}$，我们又称这类算法为非中心化(decentralized)算法：

$$x_{n+1}^1=x_n^1+a_ny_{n+1}^1=x_n^1+a_n(h(x_n^1)+\varepsilon_{n+1}^1), \qquad (8.4.1a)$$

$$\cdots$$

$$x_{n+1}^r=x_n^r+a_ny_{n+1}^r=x_n^r+a_n(h(x_n^r)+\varepsilon_{n+1}^r), \qquad (8.4.1b)$$

$$z_{n+1}=\sum_{i=1}^r\lambda_{n+1}^ix_{n+1}^i,\ \sum_{i=1}^r\lambda_n^i=I,\ \forall n. \qquad (8.4.1c)$$

其中 y_{n+1}^i 表示第 i 个观测器在时刻 n 的观测量，ε_{n+1}^i 为此时的观测噪声，$\{a_n\}$为共同的步长因子序列，$\{\lambda_n^i\}$是矩阵列，满足 $\sum_{i=1}^{r}\lambda_n^i=I$，$\forall n$.

采用式(3.1.3)和(3.1.4)处的记号，我们还可将上述算法中的(8.4.1a)和(8.4.1b)改造成它们的变界截尾形式：对 $i=1, 2, \cdots, r$,

$$\sigma_0^i=0,\ \sigma_n^i=\sum_{j=0}^{n-1} I_{[\|x_j^i+a_j y_{j+1}^i\|>M_{\sigma_j^i}]}, \tag{8.4.2}$$

$$x_{n+1}^i=(x_n^i+a_n y_{n+1}^i) I_{[\|x_n^i+a_n y_{n+1}^i\|\leqslant M_{\sigma_n^i}]}+x^* I_{[\|x_n^i+a_n y_{n+1}^i\|>M_{\sigma_n^i}]}, \tag{8.4.3}$$

于是式(8.4.2)、(8.4.3)和(8.4.1c)为算法(8.4.1)的变界截尾形式.

算法 II

和新药最佳剂量确定的问题不同，空中飞行器最佳参数的调控不能由在地面上的多个处理器各自独立地进行，因为它们调控的是同一个对象，那样作会彼此干扰．在任何时刻 n，所有的观测器都只能取同样的观测点(这里也是被调控的参数)，但可以利用各自得到的观测进行一步迭代．显然，各个处理器得到的迭代值是不一样的，我们凸组合这些值作为下一次迭代处理的共同观测点．我们又称这类算法为中心化(centralized)算法，用数学公式叙述如下：

$$x_{n+1}^i=z_n+a_n(h(z_n)+\varepsilon_{n+1}^i), \quad 1\leqslant i\leqslant r, \tag{8.4.4a}$$

$$z_n=\sum_{i=1}^{r}\lambda_n^i x_n^i,\ \sum_{i=1}^{r}\lambda_n^i=I,\ \forall n. \tag{8.4.4b}$$

类似，算法(8.4.4)也有如下相应的变界截尾形式：

$$\sigma_0=0,\ \sigma_{n+1}=\sum_{j=0}^{n} I_{[\|\sum_{i=1}^{r}\lambda_j^i x_j^i\|>M_{\sigma_j}]}, \tag{8.4.5a}$$

$$x_{n+1}^i=z_n+a_n(h(z_n)+\varepsilon_{n+1}^i), \quad 1\leqslant i\leqslant r, \tag{8.4.5b}$$

$$z_n=\left(\sum_{i=1}^{r}\lambda_n^i x_n^i\right) I_{[\|\sum_{i=1}^{r}\lambda_n^i x_n^i\|\leqslant M_{\sigma_n}]}+x^* I_{[\|\sum_{i=1}^{r}\lambda_n^i x_n^i\|>M_{\sigma_n}]},$$

$$\sum_{i=1}^{r}\lambda_n^i=I,\ \forall n. \tag{8.4.5c}$$

从前面的第3、4章，我们已经可以看到，在讨论随机逼近算法的a.s.收敛性和渐近正态性时，变界截尾的形式正好取代了预先假设标准算法一致有界，其余的假设两种算法都相同，因此为了不重复，我们在下面只讨论和分析算法I中的算法(8.4.1)和算法II中的算法(8.4.4)，所得结果同样适用于它们的变界截尾形式.

由于算法(8.4.1)中的$\{x_n^i\}\ i=1,\ \cdots,\ r$实际上就是单处理器时的标准随机逼近算法，如果再假设对每一权系数阵序列$\{\lambda_n^i\}$，都有$\lambda_n^i\xrightarrow[n]{}\lambda^i$，则由$x_n^i$的a.s.收敛性不难得到由(8.4.1c)定义的$z_n$的a.s.收敛性，而对算法(8.4.4)，由于容易重新写它为

$$z_{n+1}=z_n+a_n\left(h(z_n)+\sum_{i=1}^{r}\lambda_{n+1}^i\varepsilon_{n+1}^i\right), \tag{8.4.6}$$

因此它的a.s.收敛性分析也成了平凡的问题. 这样，对两类算法我们都不再讨论a.s.收敛性问题，而是把注意力集中在如何在最小方差的意义下，分析出最优凸组合系数λ^{i*}，以及如何构造出$\{\lambda_n^i\}$，使得$\lambda_n^i\to\lambda^{i*}$.

设每个观测器的观测噪声分为共同部分$\{e_n\}$和彼此独立的部分$\{\xi_n^i\}$，$i\leqslant r$，即

$$\varepsilon_n^i=e_n+\xi_n^i, \tag{8.4.7}$$

$$\mathrm{E}(e_n|\mathscr{F}_{n-1})=\mathrm{E}(\xi_n^i|\mathscr{F}_{n-1})=0,\quad i\leqslant r, \tag{8.4.8}$$

$$\mathrm{E}(e_ne_n^{\mathrm{T}}|\mathscr{F}_{n-1})\xrightarrow[n]{}\sigma\ \text{a.s.},\ \mathrm{E}(\xi_n^i\xi_n^{i\mathrm{T}}|\mathscr{F}_{n-1})\xrightarrow[n]{}R^i\ \text{a.s.},\ i\leqslant r, \tag{8.4.9}$$

$$\mathrm{E}(e_n\xi_n^{i\mathrm{T}}|\mathscr{F}_{n-1})=\mathrm{E}(\xi_n^i\xi_n^{j\mathrm{T}}|\mathscr{F}_{n-1})=0,\ \forall i:i\leqslant r,\ i\neq j, \tag{8.4.10}$$

$$\sup_n\mathrm{E}(\|e_n\|^\delta+\|\xi_n^i\|^\delta|\mathscr{F}_{n-1})<\infty,\text{对某一}\ \delta>2. \tag{8.4.11}$$

注8.4.1 在两类算法的实际背景中，观测噪声可以如式(8.4.7)那样表示成两部分的和是合理的，例如新药制造过程中的

干扰和飞行器在飞行环境中遇到的干扰都可以看作各观测噪声中的共同部份 $\{e_n\}$，而各个医院在试验中的干扰和飞行器调控中各观测仪器的误差，都可以视作相互独立的观测噪声部分 $\{\xi_n^i\}$．为了突出本节关心的凸组合优化问题，避免描述上的复杂性，我们仅假设 $\{\xi_n^i\}$ 为简单的鞅差序列，但这里的结果可以容易地推广到 $\{\xi_n^i\}$ 为如 A4.1.3 那样依时间无穷相关的随机序列．

我们先给出算法的渐近正态性定理．

定理 8.4.1　设观测噪声序列满足式(8.4.7)～(8.4.11)，回归函数 $h(\cdot)$ 满足条件 A4.2.0，步长因子 $\{a_n\}$ 满足 A4.2.1，又设算法(8.4.1)中 $x_n^i \xrightarrow[n]{} x^0$ a.s.，$i\leqslant r$，算法(8.4.4)中 $z_n \xrightarrow[n]{} x^0$ a.s.，而且均有 $\lambda_n^i \xrightarrow[n]{} \lambda^i$　a.s.，$i\leqslant r$，λ_n^i 是 $\mathscr{F}_{n-1}$ 可测．那么对算法(8.4.1)有

$$\frac{z_n-x^0}{\sqrt{a_n}} \xrightarrow[n\to\infty]{d} N(0,\ S_1), \tag{8.4.12}$$

$$S_1=\int_0^\infty \left(e^{(H+\frac{1}{2}\alpha I)t}\sigma e^{(H+\frac{1}{2}\alpha I)^\tau t}+\sum_{i=1}^r \lambda^i e^{(H+\frac{1}{2}\alpha I)t}R^i e^{(H+\frac{1}{2}\alpha I)^\tau t}\lambda^{i\tau}\right)\mathrm{d}t, \tag{8.4.13}$$

而对算法(8.4.4)有

$$\frac{z_n-x^0}{\sqrt{a_n}} \xrightarrow[n\to\infty]{d} N(0,\ S_2), \tag{8.4.14}$$

$$S_2=\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}\left(\sigma+\sum_{i=1}^r \lambda^i R^i \lambda^{i\tau}\right)e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t. \tag{8.4.15}$$

证明　对算法(8.4.1)，我们有

$$\begin{aligned}(z_n-x^0)/\sqrt{a_n}&=\sum_{i=1}^r \lambda_n^i(x_n^i-x^0)/\sqrt{a_n}\\&=\sum_{i=1}^r \lambda^i(x_n^i-x^0)/\sqrt{a_n}+\sum_{i=1}^r(\lambda_n^i-\lambda^i)(x_n^i-x^0)/\sqrt{a_n}.\end{aligned} \tag{8.4.16}$$

由定理的假设和类似定理 4.2.1 的推导，知上式右端第一项可写成

$$\sum_{i=1}^r \lambda^i(x_n^i-x^0)/\sqrt{a_n}=\sum_{i=1}^r \lambda^i\sum_{k=1}^n \Psi_{n,k+1}\frac{a_k}{\sqrt{a_{k+1}}}\varepsilon_{k+1}^i+o(1)$$

$$= \sum_{k=1}^{n} \Psi_{n,k+1} \frac{a_k}{\sqrt{a_{k+1}}} e_{k+1} + \sum_{i=1}^{r} \lambda^i \sum_{k=1}^{n} \Psi_{n,k+1}$$

$$\cdot \frac{a_k}{\sqrt{a_{k+1}}} \xi_{k+1}^i + o(1), \tag{8.4.17}$$

其中记号 $\Psi_{n,k+1}$ 如§4.1和§4.2中定义, $o(1) \xrightarrow[n]{p} 0$

再由式(8.4.7)～(8.4.11)、(8.4.17)，则可以把(8.4.17)中的 $\Psi_{n,k+1} \frac{a_k}{\sqrt{a_{k+1}}} e_{k+1} + \sum_{i=1}^{r} \lambda_i \Psi_{n,k+1} \frac{a_k}{\sqrt{a_{k+1}}} \xi_{k+1}^i$ 看作引理4.1.1中的 ξ_{nk}，且能验证对这样的 ξ_{nk} 该引理的条件均成立，再注意 e_k 与 ξ_k^i 及 ξ_k^j 之间的无关性，知渐近方差为 S_1，此即

$$\sum_{i=1}^{r} \lambda^i (x_n^i - x^0) / \sqrt{a_n} \xrightarrow[n\to\infty]{d} N(0, S_1) \tag{8.4.18}$$

而由 $\lambda_n^i \xrightarrow[n]{} \lambda^i$ 知，式(8.4.16)右端的第二项以概率趋于0，故定理结论中的式(8.4.12)和式(8.4.13)成立.

对算法(8.4.4)，注意由式(8.4.9)、(8.4.10)和 λ_n^i 是 $\mathscr{F}_{n-1}$ 可测，

$$\mathrm{E}\left[\left(\sum_{i=1}^{r} \lambda_n^i \varepsilon_n^i\right)\left(\sum_{i=1}^{r} \lambda_n^i \varepsilon_n^i\right)^{\tau} \Big| \mathscr{F}_{n-1}\right]$$

$$= \mathrm{E}(e_n e_n^{\tau} | \mathscr{F}_{n-1}) + \sum_{i=1}^{r} \lambda_n^i \mathrm{E}(\xi_n^i \xi_n^{i\tau} | \mathscr{F}_{n-1}) \lambda_n^{i\tau} \xrightarrow[n\to\infty]{} \sigma$$

$$+ \sum_{i=1}^{r} \lambda^i R^i \lambda^{i\tau} \quad \text{a.s.}, \tag{8.4.19}$$

于是，再由定理的其它假设，类似定理4.2.1的证明，可得定理的结论，即式(8.4.14)和(8.4.15)也成立. □

由于可以利用已有的算法渐近正态性的研究方法和结果容易地得到新算法的类似结果，所以定理8.4.1并不是本节最关心的问题. 但是有了算法的渐近正态性，以极小化渐近协方差阵 S_1 和 S_2 的迹为性能指标，就可以找到最优的凸组合系数阵 λ^{i*}. 于是我们有下面的定理.

定理8.4.2 设定理8.4.1的条件均成立，且式(8.4.9)中的 R^i, $\forall i \leqslant r$ 均可逆，则对算法(8.4.1)和(8.4.4)，分别极小化 S_1 和

S_2 的迹的最优凸组合加权阵为:

对算法(8.4.1),

$$\lambda^{i*}=\left(\sum_{j=1}^{r}(S^j)^{-1}\right)^{-1}(S^i)^{-1},\quad i\leqslant r,\qquad(8.4.20)$$

其中

$$S^i=\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}R^i e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t,\qquad(8.4.21)$$

对算法(8.4.4)

$$\lambda^{i*}=\left(\sum_{j=1}^{r}(R^j)^{-1}\right)^{-1}(R^i)^{-1},\quad i\leqslant r.\qquad(8.4.22)$$

证明 由 $\mathrm{tr}\,S_1$ 和 $\mathrm{tr}\,S_2$ 分别为形如 $F\left(\sum_{i=1}^{r}\lambda^i S^i\lambda^{i\tau}\right)$ 和 $F\left(\sum_{i=1}^{r}\lambda^i R^i\lambda^{i\tau}\right)$ 的函数, 且 S^i 和 R^i 均可逆, 于是由定理 8.3.1, 式(8.4.20)和(8.4.22)中的 λ^{i*} 为在约束 $\sum_{i=1}^{r}\lambda^{i*}=I$ 下, 上面两个函数 $\mathrm{tr}\,S_1$ 和 $\mathrm{tr}\,S_2$ 的一组极值. 此外, 容易验证, 把式(8.4.20)中的 λ^{i*} 代入式(8.4.13), 得到相应协方差阵 S_1^* 满足

$$\begin{aligned}S_1^*&=\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}\sigma e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t+\left(\sum_{i=1}^{r}(S^i)^{-1}\right)^{-1}\\&<\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}\sigma e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t+S^i,\ \forall i\leqslant r,\end{aligned}\qquad(8.4.23)$$

而把式(7.5.22)中的 λ^{i*} 代入式(8.4.15), 得到相应的 S_2^* 满足

$$\begin{aligned}S_2^*&=\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}\left(\sigma+\left(\sum_{i=1}^{r}(R^i)^{-1}\right)^{-1}\right)e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t\\&<\int_0^\infty e^{(H+\frac{1}{2}\alpha I)t}(\sigma+R^i)e^{(H+\frac{1}{2}\alpha I)^\tau t}\mathrm{d}t,\quad\forall i\leqslant r,\end{aligned}\qquad(8.4.24)$$

这也就说明 λ^{i*}, $i\leqslant r$ 确实为极小值. □

定理 8.4.2 给出了两类不同多观测器算法的最优凸组合系数, 显然, 这些系数中的 H 阵和 R^i 阵在随机逼近问题中一般是未知的, 我们不能直接应用定理 8.4.2 来实现最优凸组合系数的算法. 但是, 由定理 8.4.1 知, 只要构造一个 $\mathscr{F}_{n-1}$ 可测的 λ_n^i 序列, 使得 $\lambda_n^i\xrightarrow[n]{\text{a.s.}}\lambda^{i*}$, $i\leqslant r$, 则算法在最小的方差迹的意义下实现了凸

组合最优. 下面,我们就来讨论$\{\lambda_n^i\}$, $i\leqslant r$ 的构造问题.

由式(8.4.20)～(8.4.22)可以看出, 上述问题也可化为构造序列$\{H_n\}$和$\{R_n^i\}$, $i\leqslant r$, 使得它们都是 H_n 和 R_n^i, $i\leqslant r$ 为 $\mathscr{F}_{n-1}$ 可测的,而且

$$H_n \xrightarrow[n]{} H \quad \text{a.s.}, \quad R_n^i \xrightarrow[n]{} R^i \quad \text{a.s.}. \tag{8.4.25}$$

对回归函数 $h(\cdot)$在零点 x^0 的梯度 $\left.\dfrac{\partial h}{\partial x}\right|_{x=x^0}=H$ 的估计, 在一些有关多维适应性随机逼近的文献,如文献[91]中已有讨论,其基本思想是构造一个观测序列的差商序列, 然后作加权平均或简单的算术平均, 这样得到的序列由于加权平均去掉了观测噪声的作用,而被观测点又都趋于零点 x^0, 因此也就收敛于 H. 具体的公式及理论分析可参阅这些文献,如文献[91]中的第4节,我们在这里不再赘述了.

至于对 R^i 的估计, 则要利用噪声的性质式(8.4.7)～(8.4.11), 分别由观测序列构造出一些算术平均序列趋于$\sigma+R^i$ 和 σ,然后得到 R^i. 下面我们就给出这些估计序列的公式及收敛性定理.

对算法(8.4.1),为估计 $\sigma+R^i$, 我们建议的算法为: 对 $\forall i\leqslant r$,

$$\begin{aligned}\phi_{n+1}^i&=\frac{1}{n-1}\sum_{j=1}^{n-1}(h(x_j^i)+\varepsilon_{j+1}^i)(h(x_j^i)+\varepsilon_{j+1}^i)^{\tau}\\&=\left(1-\frac{1}{n-1}\right)\phi_n^i+\frac{1}{n-1}(h(x_{n-1}^i)+\varepsilon_n^i)(h(x_{n-1}^i)+\varepsilon_n^i)^{\tau}.\end{aligned} \tag{8.4.26}$$

为估计 σ,算法为

$$\psi_{n+1}=\left(1-\frac{1}{n-1}\right)\psi_n+\frac{1}{n-1}(h(x_{n-1}^1)+\varepsilon_n^1)(h(x_{n-1}^2)+\varepsilon_n^2)^{\tau}. \tag{8.4.27}$$

其中的上角标1, 2也可换成任意的 $i\neq j$, $i\leqslant r$, $j\leqslant r$. 显然, ϕ_n^i 和 ψ_n 都是 $\mathscr{F}_{n-1}$ 可测的.

与算法(8.4.4)类似的算法有

$$\phi_{n+1}^{i}=\left(1-\frac{1}{n-1}\right)\phi_n^i+\frac{1}{n-1}\left(h(z_{n-1})+\varepsilon_n^i\right)\left(h(z_{n-1})+\varepsilon_n^i\right)^{\tau}, \tag{8.4.28}$$

$$\psi_{n+1}=\left(1-\frac{1}{n-1}\right)\psi_n+\frac{1}{n-1}\left(h(z_{n-1})+\varepsilon_n^1\right)\left(h(z_{n-1})+\varepsilon_n^2\right)^{\tau}. \tag{8.4.29}$$

下面的定理给出了保证算法(8.4.26)～(8.4.29)具有我们所需要的收敛性的一组条件.

定理 8.4.3　设算法(8.4.1)中的 $x_n^i \xrightarrow[n]{} x^0$ a.s., $i\leqslant r$, 算法(8.4.4)中的 $z_n \xrightarrow[n]{} x^0$ a.s., 且量测噪声满足条件(8.4.7)～(8.4.11). 则式(8.4.26)和式(8.4.28)所定义的$\{\phi_n^i\}$, $i\leqslant r$ 和式(8.4.27)与式(8.4.29)所定义的$\{\psi_n\}$都具有强一致性, 即

$$\phi_n^i \xrightarrow[n]{} \sigma+R^i \quad \text{a.s.},$$

$$\psi_n \xrightarrow[n]{} \sigma \quad \text{a.s.}.$$

由于算法(8.4.26)～(8.4.29)可以看作一种线性随机递推算法, 其收敛性容易用引理 3.2.1 处理, 我们在这里不作详细证明, 有兴趣的读者可参阅文献[107]中定理 5 的证明.

若记

$$R_n^i=\phi_n^i-\psi_n,\quad A_n=H_n+\frac{1}{2}\alpha I,$$

按公式(8.4.20)～(8.4.22), 总可以得到最优凸组合加权阵的估计序列 $\{\lambda_n^{i*}\}$, $i\leqslant r$. 要注意的是在实际计算时有可能对某个 n, $\det R_n^i=0$, 某 $i\leqslant r$, 或 $A_n\leqslant 0$, 这时则令 $\lambda_n^{i*}=\lambda_{n-1}^{i*}$, $\forall i\leqslant r$. 此外, 由公式(8.4.21)计算 S_n^i 也可利用 Kronecker 积化为解线性方程:

$$(I_l\otimes A_n+A_n\otimes I_l)\operatorname{Vec} S_n^i=-\operatorname{Vec} R_n^i.$$

当 $A_n>0$, 上面方程一定有唯一解 S_n^i(参阅文献[45]中 pp31, 38, 39), 且当 $R_n^i>0$, 由 Lyapunov 引理知, $S_n^i>0$.

§8.5　多传感器的最优凸组合适应性滤波算法

在适应性滤波算法应用的实际问题中，也常常会发生在多个不同观测点用多传感器获得同一系统的多路信号的情况. 例如在用适应性滤波算法估计一个线性系统的参数时，系统的输出就很可能被人们在多个地点同时观测，这样并行得到的观测数据在适应性滤波算法中就成了并行的参考信号$\{\psi_n^i\}$，$i\leqslant p$. 又如在用适应性滤波算法作雷达、声纳和通讯中的信号滤波，实际上也大多是有多个天线和声音接收器在不同地方同时接收同一信号源发出的信号，从而得到一组并行的输出信号$\{y_n^j\}$，$j\leqslant q$. 而且，对这一类用多传感器获得同一系统的多路信号，一般可以将它们分解为两部分：

$$y_n^j=y_n+\eta_n^j,\ \psi_n^i=\psi_n+\xi_n^i,\quad i\leqslant p,\ j\leqslant q. \tag{8.5.1}$$

其中$\{y_n\}$，$\{\psi_n\}$表示各传感器所获得的共同的信号，包括所受到的共同的噪声，$\{\eta_n^j\}$和$\{\xi_n^i\}$表示各传感器所受到的相互独立的噪声，它们都是平稳的，即有

$$\mathrm{E}y_n\eta_n^{j\tau}=0,\ \mathrm{E}\psi_n\xi_n^{i\tau}=0,\ \mathrm{E}y_n\xi_n^{i\tau}=\mathrm{E}\eta_n^j\psi_n^\tau=0,\ i\leqslant p,\ j\leqslant q. \tag{8.5.2}$$

$$\mathrm{E}\eta_n^i\eta_n^{j\tau}=0,\ \mathrm{E}\xi_n^i\xi_n^{j\tau}=0,\ i\neq j;\ \mathrm{E}\eta_n^j\xi_n^{i\tau}=0,\ i\leqslant p,\ j\leqslant q. \tag{8.5.3}$$

$$\mathrm{E}y_ny_n^\tau=A,\ \mathrm{E}y_n\psi_n^\tau=B,\ \mathrm{E}\psi_n\psi_n^\tau=D,\ A>0. \tag{8.5.4}$$

$$\mathrm{E}\eta_n^j\eta_n^{j\tau}=A_j,\ \mathrm{E}\xi_n^i\xi_n^{i\tau}=D_i,\ A_j>0,\ D_i>0,\ i\leqslant p,\ j\leqslant q. \tag{8.5.5}$$

如果，我们仍采用传统的办法来处理这种情况，即把这些多个传感器合并成一个，则可重新记

$$\widetilde{y}_n=(y_n^{1\tau}y_n^{2\tau}\cdots y_n^{q\tau})^\tau,\ \widetilde{\psi}_n=(\psi_n^{1\tau}\psi_n^{2\tau}\cdots\psi_n^{p\tau})^\tau. \tag{8.5.6}$$

用式(7.1.9)、(8.5.1)～(8.5.5)，则有

$$\min_{\widetilde{x}}\mathrm{E}\|\widetilde{x}^\tau\widetilde{y}_n-\widetilde{\psi}_n\|^2=\mathrm{tr}(F-G^\tau Q^{-1}G), \tag{8.5.7}$$

其中

$$F=\mathrm{E}\tilde{\psi}_n\tilde{\psi}_n^{\tau}=\begin{pmatrix} D & \cdots & D \\ \vdots & & \vdots \\ D & \cdots & D\end{pmatrix}+\begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_p\end{pmatrix}, \tag{8.5.8}$$

$$G=\mathrm{E}\tilde{y}_n\tilde{\psi}_n^{\tau}=\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix}B(I\cdots I), \tag{8.5.9}$$

$$Q=\mathrm{E}\tilde{y}_n\tilde{y}_n^{\tau}=\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix}A(I\cdots I)+\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_q\end{pmatrix}. \tag{8.5.10}$$

利用熟知的矩阵逆的恒等式(如参考文献[22]):

$$(L+MN^{-1}M^{\tau})^{-1}=L^{-1}-L^{-1}M(N+M^{\tau}L^{-1}M)^{-1}M^{\tau}L^{-1}. \tag{8.5.11}$$

可得

$$\begin{aligned}(I\cdots I)Q^{-1}\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix} &= (I\cdots I)\left(\begin{pmatrix} A_1^{-1} & & 0 \\ & \ddots & \\ 0 & & A_q^{-1}\end{pmatrix}\right. \\ &\quad -\begin{pmatrix} A_1^{-1} & & 0 \\ & \ddots & \\ 0 & & A_q^{-1}\end{pmatrix}\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix}\left(A^{-1}+(I\cdots I)\begin{pmatrix} A_1^{-1} & & 0 \\ & \ddots & \\ 0 & & A_q^{-1}\end{pmatrix}\right. \\ &\quad \left.\left.\cdot\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix}\right)^{-1}(I\cdots I)\begin{pmatrix} A_1^{-1} & & 0 \\ & \ddots & \\ 0 & & A_q^{-1}\end{pmatrix}\right)\begin{pmatrix} I \\ \vdots \\ I\end{pmatrix} \\ &= \sum_{j=1}^{q}A_j^{-1}-\sum_{i=1}^{q}A_i^{-1}\left(A^{-1}+\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\sum_{j=1}^{q}A_j^{-1} \\ &= \left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}. \end{aligned} \tag{8.5.12}$$

于是再由式(8.5.7)～(8.5.10)和(8.5.12),有

$$\begin{aligned}\mathrm{tr}(F-G^{\tau}Q^{-1}G) &= \mathrm{tr}\sum_{i=1}^{p}\left[D+D_i-B^{\tau}\left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}B\right] \\ &\geqslant \mathrm{tr}\left[D+\left(\sum_{i=1}^{p}D_i^{-1}\right)^{-1}-B^{\tau}\left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}B\right], \end{aligned} \tag{8.5.13}$$

其中等号成立仅当 $p=1$,而不等号成立是因为对 $\forall i\leqslant p$,

$$D+D_i-B^{\tau}\left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}B$$

$$>D+\left(\sum_{i=1}^{p}D_i^{-1}\right)^{-1}-B^{\tau}\left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}B,$$

且由后面的(8.5.21)中的等式，知上式右端为半正定的阵.

另一方面，由式(8.1.13)知，此时最优滤波阵为：

$$\tilde{x}^*=(E\tilde{y}_n\tilde{y}_n^{\tau})^{-1}\mathrm{E}\tilde{y}_n\tilde{\psi}_n^{\tau},$$

当 q 和 p 很大时，由式(8.5.6)知 $\tilde{x}^*$ 的维数将变得很大，为单传感器时的 $p\times q$ 倍，这当然是我们不希望的.

在下面，我们采用的是将并行得到的输出信号和参考信号分别凸组合加权的方法，也即

$$\bar{y}_n=\sum_{j=1}^{q}\lambda_j y_n^j,\ \sum_{j=1}^{q}\lambda_j=I;\ \bar{\psi}_n=\sum_{i=1}^{p}\mu_i\psi_n^i,\ \sum_{i=1}^{p}\mu_i=I,\tag{8.5.14}$$

其中 $y_n^j\in R^{r\times l}$, $\psi_n^i\in R^{m\times l}$, $\lambda_j\in R^{r\times r}$, $\mu_i\in R^{m\times m}$, $i\leqslant p$, $j\leqslant q$. 然后在类似于式 (8.5.7) 中最小方差迹的意义下求出一组最优的凸组合系数阵$\{\lambda_j^*\}$和$\{\mu_i^*\}$，并将证明这种最优凸组合后的最小方差迹竟达到了式(8.5.13)右端的下界，然而这时的最优滤波阵 x^* 的维数为$(r\times m)$，仅为 $\tilde{x}^*$ 的 $\frac{1}{p\times q}$，始终与单传感器时的一样，这就是新方法显著的优点. 最后，与上一节类似，这样得到的最优凸组合系数阵会依赖于信号的某些未知的统计量，我们将利用信号序列给出这些未知量的递推估计方法，使这样的最优凸组合适应性滤波算法是可以实现的.

根据式(8.5.1)～(8.5.5)和(8.5.14)，有

$$\begin{aligned}\bar{A}&=\mathrm{E}\bar{y}_n\bar{y}_n^{\tau}=\mathrm{E}\left(\sum_{j=1}^{q}\lambda_j(y_n+\eta_n^j)\right)\left(\sum_{j=1}^{q}\lambda_j(y_n+\eta_n^j)\right)^{\tau}\\&=\mathrm{E}\left(y_n+\sum_{j=1}^{q}\lambda_j\eta_n^j\right)\left(y_n+\sum_{j=1}^{q}\lambda_j\eta_n^j\right)^{\tau}\\&=\mathrm{E}y_ny_n^{\tau}+\mathrm{E}\sum_{j=1}^{q}\lambda_j\eta_n^j\eta_n^{j\tau}\lambda_j^{\tau}=A+\sum_{j=1}^{q}\lambda_jA_j\lambda_j^{\tau}.\end{aligned}\tag{8.5.15}$$

$$\bar{B}=\mathrm{E}\bar{y}_n\bar{\psi}_n^{\tau}=\mathrm{E}\left(y_n+\sum_{j=1}^{q}\lambda_j\eta_n^j\right)\left(\psi_n+\sum_{i=1}^{p}\mu_i\xi_n^i\right)^{\tau}$$

$$= \mathrm{E} y_n \psi_n^{\tau} = B, \tag{8.5.16}$$

$$\overline{D} = \mathrm{E} \overline{\psi}_n \overline{\psi}_n^{\tau} = \mathrm{E}\left(\psi_n + \sum_{i=1}^{p} \mu_i \xi_n^i\right)\left(\psi_n + \sum_{i=1}^{p} \mu_i \xi_n^i\right)^{\tau}$$

$$= D + \sum_{i=1}^{p} \mu_i D_i \mu_i^{\tau}. \tag{8.5.17}$$

仍用式(7.1.9)、式(8.5.15)～(8.5.17)，我们有

$$\min_{x} E\|x^{\tau} \overline{y}_n - \overline{\psi}_n\|^2 = \operatorname{tr}(\overline{D} - B^{\tau} \overline{A}^{+} B)$$

$$= \operatorname{tr}\left(D + \sum_{i=1}^{p} \mu_i D_i \mu_i^{\tau} - B^{\tau}\left(A + \sum_{j=1}^{q} \lambda_j A_j \lambda_j^{\tau}\right)^{+} B \right). \tag{8.5.18}$$

进而，我们要求的最优凸组合加权阵$\{\mu_i^*\}$和$\{\lambda_j^*\}$应满足

$$\operatorname{tr}\left(\sum_{i=1}^{p} \mu_i^* D_i \mu_i^{*\tau} - B^{\tau}\left(A + \sum_{j=1}^{q} \lambda_j^* A_j \lambda_j^{*\tau}\right)^{-1} B \right)$$

$$= \min_{\{\lambda_j\},\{\mu_i\}} \operatorname{tr}\left(\sum_{i=1}^{p} \mu_i D_i \mu_i^{\tau} - B^{\tau}\left(A + \sum_{j=1}^{q} \lambda_j A_j \lambda_j^{\tau}\right)^{-1} B \right), \tag{8.5.19}$$

其中：$\sum_{j=1}^{q} \lambda_j = I$, $\sum_{i=1}^{p} \mu_i = I$.

利用定理 8.3.1，我们不加证明可得到下面的定理.

定理8.5.1　设信号满足式(8.5.1)～(8.5.5)，则在式(8.5.18)和(8.5.19)所确定的最小均方意义下最优凸组合系数阵为

$$\lambda_j^* = \left(\sum_{l=1}^{q} A_l^{-1}\right)^{-1} A_j^{-1}, \quad \mu_i^* = \left(\sum_{l=1}^{p} D_l^{-1}\right)^{-1} D_i^{-1}, \quad i \leqslant p, \ j \leqslant q. \tag{8.5.20}$$

这时，容易验证

$$\min_{x} \mathrm{E}\|x^{\tau} \overline{y}_n - \overline{\psi}_n\|^2$$

$$= \operatorname{tr}\left(D + \left(\sum_{i=1}^{p} D_i^{-1}\right)^{-1} - B^{\tau}\left(A + \left(\sum_{j=1}^{q} A_j^{-1}\right)^{-1}\right)^{-1} B \right)$$

$$\leqslant \operatorname{tr}(D + D_i - B^{\tau}(A + A_j)^{-1} B), \ \forall i \leqslant p, \ j \leqslant q, \tag{8.5.21}$$

$$x^* = \overline{A}^{-1} \overline{B} = \left(A + \left(\sum_{j=1}^{q} A_j^{-1}\right)^{-1}\right)^{-1} B, \tag{8.5.22}$$

定理8.5.1说明，最优凸组合权系数仅仅依赖于信号中相互独立部分的协方差$\{D_i\}$和$\{A_j\}$, $i\leqslant p$, $j\leqslant q$. 如果这些都是预先已知的，我们立即可按传统的办法构造适应性滤波算法：

$$x_{n+1}=x_n-\frac{a}{n}\bar{y}_n(\bar{y}_n^{\tau}x_n-\bar{\psi}_n^{\tau}),\ (a>0), \tag{8.5.23}$$

并且分析算法$\{x_n\}$为x^*的收敛性质（如我们在§7.2和§7.3中所做）、但是，通常人们并不预先知道D_i和A_j，因而不能直接将λ_j^*和μ_i^*代入式(8.5.14)，在算法(8.5.23)中使用. 为了获得可以实现的a.s.收敛于x^*的算法，我们利用信号$\{y_n^j\}$, $\{\psi_n^i\}$, $i\leqslant p$, $j\leqslant q$的平稳性和统计特性，式(8.5.1)～(8.5.5)可以容易地得到$\{\lambda_j^*\}$和$\{\mu_i^*\}$的估计序列$\{\lambda_j^*(n)\}$和$\{\mu_i^*(n)\}$, $i\leqslant p$, $j\leqslant q$. 再将估计序列代入式(8.5.14)，得到

$$\bar{y}_n=\sum_{j=1}^{q}\lambda_j^*(n)y_n^j,\quad \sum_{j=1}^{q}\lambda_j^*(n)=I;$$

$$\bar{\psi}_n=\sum_{i=1}^{p}\mu_i(n)\psi_n^i,\quad \sum_{i=1}^{p}\mu_i(n)=I. \tag{8.5.24}$$

最后将式(8.5.24)中的$\{\bar{y}_n\}$和$\{\bar{\psi}_n\}$带入算法(8.5.23)，就得到可实现的最优凸组合适应性滤波算法

$$x_{n+1}=x_n-\frac{a}{n}\sum_{j=1}^{q}\lambda_j^*(n)y_n^j\left(\sum_{j=1}^{q}y_n^{j\tau}\lambda_j^{*\tau}(n)x_n-\sum_{i=1}^{p}\psi_n^{i\tau}\mu_i^{*\tau}(n)\right), \tag{8.5.25}$$

其中$\lambda_j^*(n)$和$\mu_i^*(n)$又可由A_j和D_i的估计式

$$A_j(n)=\frac{1}{n}\sum_{k=1}^{n}(y_k^j-y_k^l)y_k^{j\tau},\ \forall j\leqslant q,\ l\neq j, \tag{8.5.26}$$

$$D_i(n)=\frac{1}{n}\sum_{k=1}^{n}(\psi_k^i-\psi_k^l)\psi_k^{i\tau},\ \forall i\leqslant p,\ l\neq i, \tag{8.5.27}$$

代入公式(8.5.20)得到：对$i\leqslant p$, $j\leqslant q$

$$\lambda_j^*(n)=\begin{cases}\left(\sum_{l=1}^{q}(A_l(n))^{-1}\right)^{-1}(A_j(n))^{-1}, & \text{当所有 } A_l(n)\text{ 可逆},\\ \lambda_j^*(n-1), & \text{否则},\end{cases} \tag{8.5.28}$$

$$\mu_i^*(n)=\begin{cases}\left(\sum\limits_{l=1}^{p}(D_l(n))^{-1}\right)^{-1}(D_i(n))^{-1}, & \text{当所有 } D_i(n) \text{ 可逆},\\ \mu_i(n-1), & \text{否则}.\end{cases}$$

$$(8.5.29)$$

为说明上述算法的收敛性，我们有下面的定理.

定理 8.5.2　设信号组$\{y_n^j\}$, $\{\psi_n^i\}$, $i\leqslant p$, $j\leqslant q$ 满足

$$\frac{1}{n}\sum_{k=1}^{n} y_k^j y_k^{j\tau} \xrightarrow[n]{\text{a.s.}} A+A_j,\ j\leqslant q, \tag{8.5.30}$$

$$\frac{1}{n}\sum_{k=1}^{n} y_k^l y_k^{j\tau} \xrightarrow[n]{\text{a.s.}} A,\ l\neq j,\ l\leqslant q,\ j\leqslant q, \tag{8.5.31}$$

$$\frac{1}{n}\sum_{k=1}^{n} y_k^j \psi_k^{i\tau} \xrightarrow[n]{\text{a.s.}} B,\ i\leqslant p,\ j\leqslant q, \tag{8.5.32}$$

$$\frac{1}{n}\sum_{k=1}^{n} (\psi_k^i-\psi_k^l)\psi_k^{i\tau} \xrightarrow[n]{\text{a.s.}} D_l,\ i\leqslant p,\ l\neq i, \tag{8.5.33}$$

$$\sup_n \frac{1}{n}\sum_{k=1}^{n}\|y_k^l y_k^{j\tau}\|<\infty \text{ a.s.},\ l\leqslant q,\ j\leqslant q, \tag{8.5.34}$$

$$\sup_n \frac{1}{n}\sum_{k=1}^{n}\|y_k^j \psi_k^{i\tau}\|<\infty \text{ a.s.},\ i\leqslant p,\ j\leqslant q. \tag{8.5.35}$$

则由式(8.5.25)～(8.5.29)给出的算法有

$$\lambda_j^*(n)\xrightarrow[n]{\text{a.s.}}\lambda_j^*,\ \mu_i^*(n)\xrightarrow[n]{\text{a.s.}}\mu_i^*,\ i\leqslant p,\ j\leqslant q, \tag{8.5.36}$$

$$x_n\xrightarrow[n]{\text{a.s.}}x^*=\left(A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}\right)^{-1}B. \tag{8.5.37}$$

证明　由算法(8.5.26)～(8.5.29)以及条件(8.5.30)、(8.5.31)和(8.5.33)，定理的结论(8.5.36)成立是显然的.

再由式(8.5.30)、(8.5.31)和(8.5.20)，知

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{l=1}^{q}\lambda_l^* y_k^l\right)\left(\sum_{j=1}^{q}\lambda_j^* y_k^j\right)^{\tau}\\&=\sum_{l=1}^{q}\lambda_l^*\sum_{j=1}^{q}\frac{1}{n}\sum_{k=1}^{n}y_k^l y_k^{j\tau}\lambda_j^{*\tau}\xrightarrow[n]{\text{a.s.}}\sum_{l=1}^{q}\lambda_l^*\sum_{j=1}^{q}A\lambda_j^{*\tau}+\sum_{j=1}^{q}\lambda_j^*A_j\lambda_j^{*\tau}\\&=A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1}.\end{aligned} \tag{8.5.38}$$

由式(8.5.32)，类似上式有

$$\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{j=1}^{q}\lambda_j^* y_k^j\right)\left(\sum_{i=1}^{p}\mu_i^*\psi_k^i\right)^{\tau}\xrightarrow[n]{\text{a.s.}} B,\tag{8.5.39}$$

于是从式(8.5.34)和(8.5.36),

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\left[\left(\sum_{l=1}^{q}\lambda_l^*(k)y_k^l\right)\left(\sum_{j=1}^{q}\lambda_j^*(k)y_k^j\right)^{\tau}-\left(\sum_{l=1}^{q}\lambda_l^* y_k^l\right)\left(\sum_{j=1}^{q}\lambda_j^* y_k^j\right)^{\tau}\right]\\&=\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{q}\sum_{j=1}^{q}(\lambda_l^*(k)y_k^l y_k^{j\tau}\lambda_j^*(k)^{\tau}-\lambda_l^*(k)y_k^l y_k^{j\tau}\lambda_j^{*\tau}\\&\quad+\lambda_l^*(k)y_k^l y_k^{j\tau}\lambda_j^{*\tau}-\lambda_l^* y_k^l y_k^{j\tau}\lambda_j^*)\\&=\frac{1}{n}\sum_{k=1}^{n}\left[\sum_{l=1}^{q}\lambda_l^*(k)\sum_{j=1}^{q}y_k^l y_k^{j\tau}(\lambda_j^*(k)-\lambda_j^*)^{\tau}\right.\\&\quad\left.+\sum_{l=1}^{q}(\lambda_l^*(k)-\lambda_l^*)\sum_{j=1}^{q}y_k^l y_k^{j\tau}\lambda_j^{*\tau}\right]\xrightarrow[n]{\text{a.s.}} 0,\end{aligned}\tag{8.5.40}$$

上式最后趋于 0 是因为式(8.6.36),

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{q}\|\lambda_l^*(k)\|\sum_{j=1}^{q}\|y_k^l y_k^{j\tau}\|\|\lambda_j^*(k)-\lambda_j^*\|\\&=\sum_{l=1}^{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{k=1}^{n}\|\lambda_l^*(k)\|\|y_k^l y_k^{j\tau}\|\|\lambda_j^*(k)-\lambda_j^*\|\\&\leqslant c\sum_{l=1}^{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{k=1}^{n}\|y_k^l y_k^{j\tau}\|\|\lambda_j^*(k)-\lambda_j^*\|,\end{aligned}\tag{8.5.41}$$

而由式(8.5.34)和(8.5.36),对 $\forall\varepsilon>0$, 存在 N_1, 当 $k\geqslant N_1$

$$\frac{1}{n}\sum_{k=N_1}^{n}\|y_k^l y_k^{j\tau}\|\|\lambda_j^*(k)-\lambda_j^*\|<\frac{\varepsilon}{2},\ l\leqslant q,\ j\leqslant q,\tag{8.5.42}$$

同时存在 N_2, 当 $n\geqslant N_2$,

$$\frac{1}{n}\sum_{k=1}^{N_1-1}\|y_k^l y_k^{j\tau}\|\|\lambda_j^*(k)-\lambda_j^*\|<\frac{\varepsilon}{2},\ l\leqslant q,\ j\leqslant q,\tag{8.5.43}$$

综合式(8.5.41)～(8.5.43),则式(8.5.40)成立.

再由式(8.5.38)和(8.5.40)可得

$$\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{l=1}^{q}\lambda_l^*(k)y_k^l\right)\left(\sum_{j=1}^{q}\lambda_j^*(k)y_k^j\right)^{\tau}\xrightarrow[n]{\text{a.s.}} A+\left(\sum_{j=1}^{q}A_j^{-1}\right)^{-1},\tag{8.5.44}$$

类似的方法,由式(8.5.35)、(8.5.36)和(8.5.39)可得

$$\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{j=1}^{q}\lambda_j^*(k)y_k^j\right)\left(\sum_{i=1}^{p}\mu_i^*(k)\psi_k^i\right)^{\tau}\xrightarrow[n]{\text{a.s.}} B.\tag{8.5.45}$$

最后，用定理7.3.1和式(8.5.44)、(8.5.45)，立即可得由算法(8.5.25)给出的$\{x_n\}$使式(8.5.37)成立。 □

至此，我们在无约束($C=0$, $\Phi=0$)的最小均方意义下，给出了一种最优凸组合适应性滤波算法，由于我们还不能放松所有A_i, D_i可逆的假设，这些结果还不能容易地推广到有约束的更一般情况。

参考文献

[1] Bar-Shalom Y. *Multitarget Multisensor Tracking.* Advances and Applications, Vol. 1, Norwood, MA: Artech, 1990

[2] Bar-shalom Y. Multitarget-Multisensor Tracking: Advances and Applications. Vol. 2, Norwood, MA: Artech, 1992

[3] Berger E. Asymptotic behaviour of a class of stochastic approximation procedures. Probab. Th. Rel. Fields, 1986, 71: 517～522

[4] Billingsley P. Convergence of Probability Measures. Wiley, 1968

[5] Blum J. Multidimensional stochastic approximation methods. Ann. Math. Stat., 1954, 25:735～744

[6] Benveniste A, Métivier M, Priouret P. Adaptive Algorithms and Stochastic Approximations. Springer Verlag, 1990

[7] Benveniste A, Ruget G. A measure of the tracking capability of recursive stochastic algorithms with constant gains. IEEE Trans, on Automatic Control, 1982, AC-27(3):639～649

[8] Bertsekas D P, Tsitsiklis J N. Parallel and Distributed Computation: Numerical Method. Prentice-Hall, 1989

[9] Chen H F Bagchi A. Parameter estimation in continuous-time dynamical systems using stochastic approximation. Technical Report

[10] Chen H F, Duncan T. and Pasik-Duncan B. On ODE approach to system parameter identification and its application to adaptive control, submitted for publication

[11] Chen H F, Duncan T and Pasik-Duncan B. Parameter identification and adaptive control of stochastic linear and semilinear systems, submitted for publication.

[12] Chen H F, Gao A T Robustness analysis of stochastic approximation. Stochastics and Stochastics Reports, 1988, 26:3～23

[13] Chen H F, Guo L, Gao A T. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. Stochastic Processes and Their Applications, 1988, 27: 217～231

[14] Chen H F, Guo L. Simultaneous estimation of both zero of regression function and parameters in noises. J. Sys. Sci &. Math. Scis., 1987, 7(2): 117～128

[15] Chen H F, Guo L. Identification and Stochastic Adaptive Control. Birkhäuser, 1991

[16] 陈翰馥. 离散时间系统的递推估计与随机控制. 科学出版社,1980

[17] Chen H F. Recursive algorithm for adaptive beam-formers. Kexue

Tongbao(Science Bulletin), 1981, 66: 490～493

[18] Chen H F. On continuous-time stochastic approximation, In: Bensoussan, Lions (Eds.), Analysis and Optimization of Systems, Lecture Notes in Control and Information Sciences. Springer-Verlag, 1982, 44: 203～214

[19] Chen H F. 相关観测误差下的随机逼近. 中国科学(A)辑, 1983, 3: 264～274

[20] Chen H F. Asymptotic normality of continuous-time stochastic approximation algorithm Acta Mathematicae Applicatae Sinica, 1984, 1(1): 31～ 3

[21] Chen H F. Asymptotic normality of stochastic approximation under correlated noise J. Sys. & Math. Scis, 1984, 4(2): 136～159

[22] Chen H F. Recursive Estimations and Control for Stochastic Systems. Wiley, 1985

[23] Chen H F. Stochastic a pproximation with randomly varying truncations in the optimization problem. Acta Math. Appl. Sinica, 1987, 10(1):58～67

[24] Chen H F. Asymptotically efficient stochastic approximation. Stochastics and Stochastics Reports, Vol. 45, 1993(112): 1～16

[25] Chen H F. Continuous time stochastic approximation: convergence and asymptotic efficiency. Stochastics and Stochastics Reports, Vol, 51, 1994, 111～132

[26] Chung K L. On a stochastic approximation method. Ann Math, Stat. 1954, 25:463～483

[27] Clark D S. Necessary and sufficient conditions for the Robbins-Monro method. Stoch. Process. Appl., 1984, 17: 359～367

[28] Chow Y S, Teicher H. Probability Theory. Springer-Verlag, 1978

[29] Chen H F, Wu C T, Zhu Y M. Continuous-time stochastic approximation Procedure with randomly varying truncations. A Mathematica Scientia, 1987, 7(1): 44～55

[30] 陈翰馥, 朱允民. 变界截尾的随机逼近算法. 中国科学(A辑), 1986, 6: 561～570

[31] 陈翰馥, 朱允民. 随机变界截尾算法的渐近性质. 数学物理学报, 1987, 7(4): 431～441

[32] Dvoresky A. On stochastic approximation. Proc. Third Berkeley Symp. On Math. Stat. and Prob., 1956, 1: 39～45

[33] Eweda E., Macchi O. Quadratic mean and almost sure convergence of unbounded stochastic approximation algorithims with correlated observations. Ann. Institut Henri Poincare 1983, 19(1)

[34] Eweda E, Macchi O. Convergence of an adaptive linear estimation algorith m IEEE Trans. on Automatic Control, 1984, AC-29(2): 119～127

[35] Eweda E, Macchi O. Convergence analysis of self-adaptive equalizers.

IEEE Trans. on Information Theory, 1984, IT-30(2): 161～176

[36] Eweda E, Macchi O. Tracking error bounds of adaptive non-stationary filtering. Automatica, 1985, 21(3): 293—302

[37] Eykhoff P. System Identification. Wiley, 1974

[38] Fabian V. On asymptotic normality in stochastic approximation. Ann. Math. Stat., 1968, 39: 1327—1332

[39] Fabian V. Stochastic approximation. In: J. Rustagi (Ed.), Optimizing Methods in Statistics, Academic Press, 1971, 439～470

[40] Fabian V. On asymptotically efficient recursive estimation. Ann. Stat., 1978, 6: 854～866

[41] Fabian V. A local asymptotic minimax optimality of an adaptive Robbins-Monro stochastic approximation procedure. In: Herenkrath U, Kalin D. Vogel W. (Eds.), Mathematical Learning Models-Theory and Algorithms, Springer Verlag, 1983

[42] Farden D. Stochastic approximation with correlated deta. IEEE Trans, on Information Theory, 1981, IT-27(1): 105～113

[43] Gelfand S B, Mitter S K. Recursive stochastic algorithms for global optimization in Rl. SIAM J. on Control and optimization, 1991, 29(5): 999～1018

[44] 龚光鲁. 随机微分方程引论. 北京大学出版社, 1987

[45] Graham A. Kronecker Products and Matrix Calculus: with Applications. Ellis Horwood Ltd, 1981

[46] Goodwin G C, Sin K. Adaptive Filtering, Prediction, and Control. Prentice Hall, 1984

[47] Hahn W. Stability of Motion. Springer-Verlag, 1967

[48] 贺建勋, 王志成. 常微分方程. 湖南科技出版社, 1981.

[49] Kushner H J, Clark D S. Stochastic Approximation for Constrained and Unconstrained Systems. Springer-Verlag, 1978

[50] Kersting G. Almost sure approximation of the Robbins-Monro process by sums of independent random variables. Ann. Probab., 1977, 5: 954～965

[51] Kushner H J, Huang H. Rates of convergence for stochastic approximation type algorithms. STAM J. Control and Opt., 1979, 17(1): 607～617

[52] Kushner H J, Huang H. Asymptotic properties of stochastic approximation with constant coefficients. SIAM J. Control and Opt., 1981, 19(1): 87～105

[53] Kushner H J, Sanvicente E. Stochastic approximation for constrained systems with observation noise on the system and constraints. Automatica, 1975, 11: 375～380

[54] Kushner H J, Shwartz A. An invariant measure approach to the convergence of stochastic approximations with state dependent noise. SIAM J. Control and Opt., 1984, 22(1): 13～27

[55] Kushner H J, Schwartz A. Weak convergence and asymptotic properties of adaptive filters with constant gains. IEEE Trans. on Information Theory, 1984, IT-30(2): 177～182

[56] Kushner H J. Stochastic approximation with discontinuous dynamics and state dependent noise with W. P. 1 and weak convergence. J. of Math. Anal. and Appl., 1981, 82: 527～542

[57] Kushner H J. Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory. MIT Press, 1984

[58] Kushner H J. An averaging method for stochastic approximation with discontinuous dynamics, constraints, and state dependent noise, Recent Advances in Statistics. A cademic Press, 1983, 211～235

[59] Kiefer J, Wolfowitz J. Stochastic estimation of the modulus of a regression function. Ann. Math. Stat., 1952, 23: 462～466

[60] Kushner H J, Yang J. Stochastic approximation with averaging of the iterates: optimal asymptotic rate of convergence for general processes. SIAM J. on Control and Optimization, 1992, 31(4): 1045～1062

[61] Kushner H J, Yang J. Stochastic approximation with averaging and feedback: rapidly convergent "on line" algorithms, and applications to adaptive systems. Tech. Report LCDS #92-8, Division of Applied Math., Brown University, 1992

[62] Kushner H J, Yin G. Asymptotic properties of distributed and communicating stochastic approximation algorithms. SIAM J. on Contr. Opt. 1987, 25: 1266～1290

[63] Kushner H J, Yin G. Stochastic approximation algorithms for parallel and distributed processing. Stochastics, 1987, 22: 219～250

[64] Lasalle J P. The Stability of Dynamic Systems. SIAM, 1976

[65] Ljung L. Analysis of recursive stochastic algorithms. IEEE Trans, on Automatic Control, 1977, AC-22(4): 551～575

[66] Ljung L. Convergence of an adaptive filter algorithm. Int. J. Control, 1978, 27(5): 673～693

[67] Ljung L. Analysis of stochastic gradient algorithms for linear regression problems. IEEE Trans. on Information Theory, 1984, IT-30:151～160

[68] Long G Z Ling LY, Proackas JG. The LMS algorithm with delayed coefficient adaptation. IEEE Trans. on Acous. Spee. Sign. Processing, 1989, ASSP-37(9):1397～1405

[69] Lai T L, Robbins H. Limit theorems for weighted sums and stochastic approximation processes. Proc. Nat. Acad. Sci. U. S. A. 1978, 75: 1068～1070

[70] Lai T L, Robbins H. Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. Wahrs., verw, Gebiete, 1981, 56: 329～360

[71] Ljung L, Soderströnf T. Theory and Practice of Recursive Identification. MIT Press, 1983

[72] Liptser R S, Shiryayev A N. Statistics of Random Processes. Springer-Verlag, 1977

[73] Lai T L, Wei C Z. Least squares estimations in stochastic regression models with applications to identification and control of dynamic systems. Ann. Stat.,1982, 10(1): 154～166

[74] Nevelson M B, Hasminskii R Z. Stochastic Approximation and Recursive Estimation. AMS Translations of Math. Monographs, 1976, 47

[75] Pflug G. Stochastic minimization with constant step size:asym-ptotic laws. SIAM J. Control and Opt., 1986, 24: 655～666

[76] Polyak B T, Juditsky A B. Accelaration of stochastic approximation by averaging. SIAM J. on Control and Optimization, 1992, 30: 838～855

[77] Polyak B T. New method of stochastic approximation type. Automat. Remote Control, 1990, 51: 937～946

[78] Robbins H, Monro S. A stochastic approximation method. Ann. Math. Stat., 1951, 22: 400～407

[79] Rodrigue G. Parallel Computations. Academic Press, 1982

[80] Ruppert D. Almost sure approximation to the Robbins-Monro and Kiefer-Wolfowitz Processes with dependent noise. Ann. Probab., 1982, 10(1): 128～137

[81] Ruppert D. Convergence of stochastic approximation algorithms with non-additive dependent disturbances and applications, In: Herenkrath U., Kalin D., Vogel W. (Eds.), Mathematical Learning Models-Theory and Algorithims. Springer-Verlag, 1983, 182～190

[82] Ruppert D. Efficient estimators from a slowly convergent Robbins-Monro process, Technical Report, School of Operations Research and Industrial Engineering, Cornell University, 1988, Tech. Report 787

[83] Sacks J. Asymptotic distributions of stochastic approximation procedures. Ann. Math Stat., 1958, 29: 373～405

[84] Solo V. Stochastic approximation with dependent noise. Stochastic Processes and Their Applications, 1982, 13: 157～170

[85] Tsitsiklis J N. Problems in decentralized decision making and

computation. *Ph. D. thesis, Dept. of Electrical Engineering* and Computer Science, MIT, Cambridge, MA, USA

[86] Venter J. An extension of the Robbins-Monro Procedure. Ann. Math. Stat., 1967, 38: 181～190

[87] Walk H. An invariance principle for the Robbins-Monro Process in a Hilbert space Z. Wahrs. verw. Gebiete, 1977, 39: 135—150

[88] Walk H. Almost sure convergence of stochastic approximation processes. Statistics and Decisions, Supplement Issue 1985, 2: 137—141

[89] Walk H. Limit behaviur of stochastic approximation processes. Statistics and Decisions, 1988, 6: 109～120

[90] Wei C Z. Asymptotic properties of Least-Squares estimates in stochastic regression models. Ann. Stat., 1985, 13(4): 1498～1508

[91] Wei C Z. Multivariate adaptive stochastic approximation. Ann. Stat., 1987, 15(3):1115～1130

[92] Widrow B, Stearns S. Adaptive Signal Processing. Prentice Hall, 1985

[93] 夏道行,吴卓人,严绍宗,舒五昌. 实变函数论与泛函分析. 高等教育出版社, 1979

[94] Yin G. Asymptotic properties of an adaptive beam former algorithm. IEEE Trans. on Information Theory, 1989, IT-35: 937～946

[95] Yin G. On extensions of polyak's averaging approach to stochastic approximation. Stochastics and Stochastics Reports, 1992, 36: 245～264

[96] Yin G. Stochastic approximation via averaging Polyak's approach revised. In: Pflug G, Dieter U. (Eds.), Lecture Notes in Economics and Mathematical Systems 374, Springer-Verlag, 1992, 119～134

[97] Yin G., Zhu Y M. Almost sure convergence of a stochastic approximation algorithm with non-additive noise. Int. J. Control, 1989, 49: 1361～1376

[98] Yin G, Zhu Y M. On W. P. 1. convergence of a parallel stochastic appoximation algorithm. Probability in the Eng. and Infor. Sciences, 1989, 3: 55～75

[99] Yin G, Zhu Y M. Averaging procedures in adaptive filtering: an efficient approach. IEEE Trans. on Automatic Control, 1992, AC-37(4): 466～475

[100] Zhu Y M, Gao A J. Necessary and sufficient condition and robustness for convergence of adaptive filtering algorithm. IEEE Proc. of the 30th CDC, 1991, 1792～1793

[101] Zhu Y M, Gao A J. Optimal design of convex combinations for adaptive filtering with parallel signals. 1994 International Symposium on Speech, Image Processing and Neural Networks Proceedings, Hong Kong, 1994, 2: 741～744

[102] 朱允民. 在相关量测误差下的一种截尾随机逼近算法. 应用数学学报, 1985, 8(1): 93～100

[103] 朱允民. 级数和无穷乘积关系的推广和一类递推算法的收敛性质. 计算数学,1985, 4: 369～376

[104] 朱允民. 适应性波束形成器的强收敛速度. 系统科学与数学, 1990, 4: 289～295

[105] Zhu, Y M. Asymptotic normality for a vector stochastic difference equation with applications in stochastic approximation. Rearch Report IMS-44, Institute of Math. Sciences, Chengdu Branch, Academia Sinica, 1992. Accepted by Journal of Multivariate Analysis.

[106] Zhu Y M, Yin G. Adaptive filter with constraint and correlated non-stationary signals. Systems and Control Letters, 1988, 4: 271～279

[107] Zhu Y M, Yin G. Optimal quasi-convex combination for stochastic approximation algorithm with parallel observations. Int. J. Control, 1989, 49: 1947～1964

[108] Zhu Y M, Yin G. Convergence speed and asymptotic distribution of a parallel Robbin-Monro method. J. of Computational Mathematics, 1990, 8: 45～54

[109] Zhu Y M, Yin G. A new algorithm for constrained adaptive array processing. Int. J. Adaptive Control and Signal Processing, 1990, 4: 259～269

[110] Zhu Y M, Yin G. Stochastic approximation in real time: A pipe line approach. J. of Computational Mathematics, 1994, 12(1): 21～30

后　记

一个作品完成后，特别在它排印后，作者往往会感到想补充或想改动而又无法再变的遗憾．近一两年来，国外学者及作者本人在随机逼近研究方面又取得一系列进展．如有机会改写本书，作者会把它们包含进去，但现在已不可能．作为一个补救，下面作一些扼要补充，但不加证明．

后记文献[1]的作者在对回归函数及步长因子加了一些条件后，论述了对量测噪声的几个条件的等价性，其中也包括本书的条件 A2.4.3，从定理 3.4.2 的证明可以看出，它们都等价于以下条件：

量测噪声 ϵ_i 可分为两部分：$\epsilon_i=\epsilon_i^{(1)}+\epsilon_i^{(2)}$ 使

$$\sum_{i=1}^{\infty} a_i\epsilon_{i+1}^{(1)}<\infty,\quad \epsilon_i^{(2)}\xrightarrow[i\to\infty]{}0. \tag{后 1}$$

但应该指出，这些等价性只有对回归函数和步长因子加了条件后才成立．在这些附加条件下，当 A2.4.3 成立时，如果 $J=x^0$，那么 $\{x_n\}$ 本身收敛，于是 A2.4.3 中的 $\{x_{n_k}\}$ 可取 $\{x_n\}$ 本身，所以 A2.4.3 等价于

$$\lim_{T\to 0}\limsup_{n\to\infty}\frac{1}{T}\left\|\sum_{i=n}^{m(n,T_k)} a_i\epsilon_{i+1}\right\|=0,\quad \forall T_k\in(0,T), \tag{后 2}$$

也就是要求它对整个序列 $\{x_n\}$ 都成立．

但实际上，这些"等价条件"本身并不等价，相比之下，以本书的 A2.4.3 最弱．因为未证定理之前，$\{x_n\}$ 本身是否有界也不清楚，当然谈不上它的收敛性．所以条件(后 2)不好验证，而条件 A2.4.3 只要验证它在收敛子列 $\{x_{n_k}\}$ 上成立，这就容易得多．这一点在第六章的定理证明中，已经看到，那里验证了条件 A2.4.3，但在证明 $\{x_n\}$ 收敛或有界前要直接验证(后 1)或(后 2)却几乎是不可能的．在后记文献 2～4]中，我们也成功地验证了 A2.4.3，若要直接验证(后 1)或(后 2)，是很难想象的．

定理 3.1.1 是分析算法收敛性的强有力工具，利用这个定理，我们解决了模式识别中主分量分析问题[2]，带随机差分的 KW 算法的收敛性问题[3]，一个适应镇定问题中的参数选择问题[4]，以及在 DEDS 中的参数优化问题[5]。在后记文献[6]中对这类工作有一个概述，在这里无法一一叙述，但后记文献[3]中的结果，对定理 3.1.4 有本质改进。

通常的 KW 算法，每算一步需要 $2l$ 个量测（见(3.1.29)），当 l 很大时，尤为不便。后记文献[8]中建议用随机差分，每算一步，只需二个量测，但为使算法收敛，后记文献[8]中所用条件十分苛刻。利用定理 3.1.1，得到随机差分下 KW 算法新的收敛定理，去掉了后记文献[8]中所用的五个限制性条件。

设 $\{\Delta_k^i,\ i=1,\ \cdots,\ l,\ k=1,\ 2,\ \cdots\}$ 是独立同分布的随机变量序列，$E\Delta_k^i=0$，$|\Delta_k^i|<a$，$E(1/\Delta_k^i)=0$，$|1/\Delta_k^i|<b$，a，b 为正常数。记 $\Delta_k=[\Delta_k^1\cdots\Delta_k^l]^\tau$，

$$g_k=\left[\frac{1}{\Delta_k^1},\ \cdots,\ \frac{1}{\Delta_k^l}\right]^\tau.$$

设 $h(\cdot)$ 为 $R^l\longrightarrow R$ 函数。在对任一时刻，取两个量测

$y_{k+1}^+=h(x_k+c_k\Delta_k)+\xi_{k+1}^+$，$y_{k+1}^-=h(x_k-c_k\Delta_k)+\xi_{k+1}^-$，$\xi_{k+1}^+$ 和 ξ_{k+1}^- 为量测噪声。

组成差分

$$\begin{aligned} y_{k+1}&=\frac{y_{k+1}^+-y_{k+1}^-}{2c_k}\,g_k \\ &=\frac{h(x_k+c_k\Delta_k)-h(x_k-c_k\Delta_k)}{2c_k}\,g_k+\frac{\xi_{k+1}}{2c_k}g_k, \end{aligned} \tag{后 3}$$

$$\xi_{k+1}=\xi_{k+1}^+-\xi_{k+1}^-. \tag{后 4}$$

定理 1 设 1) $f(x)\triangleq h_x(x)$ 满足局部 Lipschitz 条件，$f(x^0)=0$，$f(x)\neq 0$，$\forall x\neq x^0$，x^0 是 $h(\cdot)$ 的极大值。2) 还设存在常数 c_0 及矢量 x^*，使 $\sup\limits_{\|x\|=c_0}h(x)<h(x^*)$。3) 又设步长因子 a_k，c_k 如下：$a_k>0$，$c_k>0$，$c_k\to 0$，$\sum\limits_{k=1}^{\infty}a_k=\infty$，且对某个 $p\in(0,\ 1]$，$\sum\limits_{k=1}^{\infty}a_k^{1+p}<\infty$，

那么若(3.1.3)及(3.1.4)中的 y_{k+1} 由(后3)给出，则 $x_n \to x^0$ a. s. 的充分必要条件是对每个 j, $j=1, \cdots, l$, 由(后4)定出的 ξ_k 可分解成

$$\xi_k = e_k^j + \nu_k^j$$

使得

$$\sum_{k=1}^{\infty} \frac{a_k e_{k+1}^j}{c_k \Delta_k^j} < \infty \quad \text{a. s.,} \quad \frac{\nu_k^j}{c_k \Delta_k^j} \xrightarrow[k\to\infty]{} 0 \quad \text{a. s.} \qquad (后5)$$

下面的定理给出使(后5)成立的充分条件.

定理2. 设定理1的条件1), 2), 3)成立, 还设$\{\xi_k\}$和$\{\Delta_k\}$相互独立, $\sum_{k=1}^{\infty} a_k^2/c_k^2 < \infty$, 并且下列两条件之一成立: i) $\sup_k |\xi_k| \leqslant \xi < \infty$ a.s. ii) $\sup_k E\xi_k^2 < \infty$, 那么由(3.1.3), (3.1.4)及(后3)定义的 $x_n \xrightarrow[n\to\infty]{} x^0$ a. s.

在§3.2中我们给出了算法的收敛速度. 从条件(3.2.1)看出 $h_x(x^0) = H \neq 0$, 称它为非退化情形. 在后记文献[7]中我们研究了退化情形: $h_x(x^0) = 0$ 时的收敛速度. 当步长因子为 $a_n = \frac{1}{n^\alpha}$, $0 < \alpha \leqslant 1$ 时, 从定理3.2.1看出, 对非退化情形, 收敛速度为 $\|x_n - x^0\| = o(n^{-\alpha\delta})$, 所以当 $\alpha = 1$ 时收敛速度优于 $\alpha \in (0, 1)$ 时的收敛速度. 但对退化情形, 情况恰恰相反, $\alpha = 1$ 时的收敛速度劣于 $\alpha \in (0, 1)$ 时的收敛速度. 这可以从下面的定理[7]可以看出.

定理3　设成立以下条件:

1) $h(\cdot)$: $R^l \to R^l$ 为局部有界, $h(x^0) = 0$, $h(x) \neq 0$, $\forall x \neq x^0$;

2) 存在二次可微函数 $v(\cdot)$: $R^l \to R$, 使

$$\sup_{\delta_1 \leqslant \|x - x^0\| \leqslant \delta_2} h^\tau(x) v_x(x) < 0, \ \forall \delta_2 > \delta_1 > 0$$

3) 当 $x \to x^0$ 时,

$$h(x) = H(x - x^0)\|x - x^0\|^\gamma + r(x), \ \gamma > 0,$$

$$r(x) \in R^l, \ r(x)/\|x - x^0\|^{1+\gamma} \longrightarrow 0,$$

H 的所有本征值都有负实部.

$$4)\ a_k>0,\ a_k \xrightarrow[k\to\infty]{} 0,\ \sum_{i=1}^{\infty} a_i=\infty,$$

$$q_n \triangleq c_{n+1}^{-1}-c_n^{-1},\quad q_n\geqslant 0,\quad \limsup_{n\to\infty} q_n=q,\ 0\leqslant q<\infty,$$

$$\sum_{i=1}^{\infty}\frac{a_i}{\log a_i^{-1}}<\infty,$$

5) 量测噪声$\{\epsilon_i\}$满足条件: $\sum_{i=1}^{\infty} a_i(\log a_i^{-1})^{\frac{1}{\gamma}}\epsilon_{i+1}<\infty$ 那么由式(3.1.3)、(3.1.4)给出的 x_n 有如下收敛速度

$$\limsup_{n\to\infty}(\log c_n^{-1})^{\frac{1}{\gamma}}|x_n-x^0|\leqslant\sqrt{k}\left(\frac{2q\lambda_{\max}}{\gamma}\right)^{1}, \qquad (后 6)$$

而对 $l=1$ 时,(后 6)的上界达到.

这里 k 和 $\lambda_{\max}$ 分别为 P 的条件数和最大本征值, 而 P 为 Lyapunov 方程 $PH+H^{\tau}P=-I$ 的解.

当 $a_n=n^{-\alpha}$ 时,

$$q=\begin{cases}0, & 若\ \alpha\in(0,\ 1)\\ 1, & 若\ \alpha=1.\end{cases}$$

所以至少在一维时, $\alpha\in(0,\ 1)$比 $\alpha=1$ 时的收敛速度好. 退化和非退化情形在收敛速度方面的相反性质, 在数字模拟上也得到了验证,不能不感叹随机逼近的奇妙之处.

陈 翰 馥

1995 年 1 月 9 日于北京

后记文献

[1] Wang I J, Chong E K P, Kulkarni S R, Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms, to appear in Advances in Applied Probability

[2] Zhang J H, Chen H F, Convergence of algorithms used for principal component analysis, submitted for publication

[3] Chen H F, Duncan T, Pasik-Duncan B, A Kiefer-Wolfowitzalgorithm with randomized differences, submitted for publication

[4] Chen H F, Cao X R, Controllability is not necessary for adaptive pole placement control, submitted for publication

[5] Tang Q Y, Chen, H F, Convergence of perturbation analysis based optimization algorithm with fixed-number of customers period, Discrete Event Dynamic Systems, No. 4, 1994, 359～373

[6] Chen, H F, Stochastic approximation and its new applications, Proceedings of 1994 Hong Kong International Workshop on New Directions of Control and Manufacturing, 1994, 2～12

[7] Chen, H F, Convergence rate of stochastic approximation algorithms in the degenerate case, submitted for publication

[8] Spall J C, Multivariance stochastic approximation using a simultaneous perturbation gradient approximation, IEEE Trans, Autom. Control, Vol. 37, No. 3, 1992, 332～341